

```

DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
DE Site-specific recombinase, phage integrase family.
GN OrderedLocustNames=500388;
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCB1_Taxid=70663;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RC MEDLINE=22297686; PubMed=12368813; DOI=10.1038/nbt749;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N.L., Melhe B.A.,
RA Clayton R.A., Meyer T., Tsapin A., Scott J., Beanan M.J.,
RA Brinkac L.M., Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S.,
RA Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A.,
RA White O., Wolf A.M., Vamathevan J.J., Weidman J.F., Impraim M.,
RA Lee K., Berry K.J., Lee C., Mueller J., Khouli H.M., Gill J.,
RA Ullrich T.R., McDonald L.A., Feldblyum T.V., Smith H.O.,
RA Venter J.C., Nealon K.H., Fraser C.M.,
RT Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015487; AAN53471.1; -.
DR TIGR; S00388; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0015074; P:DNA integration; IEA.
DR GO; GO:0006310; P:DNA recombination; IEA.
DR InterPro; IPR011010; DNA_birk_join_enz.
DR InterPro; IPR002104; Phage_integrase.
DR Pfam; PF00589; Phage_integrase; 1.
KW Complete proteome.
SQ
SEQUENCE 429 AA; 48452 MW; E6AC2F3ED4E113D CRC64;

Query Match
Best Local Similarity 94.3%; Score 33; DB 2; Length 429;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7
DB 294 DSGELKR 300

RESULT 3
Q9PP43 PRELIMINARY; PRT; 643 AA.
AC Q9PP43;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE ABC transport system ATP-binding protein.
GN OrderedLocustNames=Cj0888.3, Cj0888c;
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC Campylobacteraceae; Campylobacter.
OX NCB1_Taxid=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RC MEDLINE=20150912; PubMed=10668204; DOI=10.1038/35001088;
RA Parkhill J., Wren B.W., Mungall K.L., Kelsey J.M., Churcher C.M.,
RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S.,
RA Jørgensen K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rastendream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.
RL Nature 403:665-668(2000).
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AL139076; CAB73146.1; -.
DR PIR; H81361; H81361.
DR HSSP; P58301; IUS8.
DR GO; GO:0016020; C:membrane; IEA.

```

```

DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0042625; F:ATPase activity; coupled to transmembrane m. .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; AAA_Transporter.
DR Pfam; PF000005; ABC_tran; 2.
DR Pfam; PF000006; ABC_transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
KW ATP-binding; Complete proteome.
SQ
SEQUENCE 643 AA; 73915 MW; F23FCA0DFE0CE80D5 CRC64;

Query Match
Best Local Similarity 94.3%; Score 33; DB 2; Length 643;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7
DB 384 DSGELKR 390

RESULT 4
RL32_BRAJA
ID RL32_BRAJA STANDARD; PRT; 60 AA.
AC Q89V07;
DT 25-OCT-2004 (rel. 45, Created)
DT 25-OCT-2004 (rel. 45, Last sequence update)
DT 25-OCT-2004 (rel. 45, Last annotation update)
DE 50S ribosomal protein L32.
GN Name-rpmf, OrderedLocustNames=br0948;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCB1_Taxid=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RC MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiyama T.,
RA Sasamoto S., Watanabe A., Idegawa K., Itiguchi M., Kawashima K.,
RA Kohara M., Matsunoto M., Shimo S., Tsurucka H., Wada T., Yamada M.,
RA Tabata S.;
RT Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.
RL DNA Res. 9:189-197(2002).
CC -1- SIMILARITY: Belongs to the ribosomal protein L32P family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC DR EMBL; AP005938; BAC46213.1; -.
CC DR HAMAP; MF_00340; -.
CC DR InterPro; IPR002677; Ribosomal_L32P.
CC DR InterPro; IPR005718; Ribosomal_L32P/o.
CC DR Pfam; PF01783; Ribosomal_L32P; 1.
CC DR TIGRPFam; TIGR01031; rpmf_bact; 1.
CC KW Complete proteome; Ribosomal protein.
SQ
SEQUENCE 60 AA; 6979 MW; 5D0F3B445AF136D9 CRC64;

Query Match
Best Local Similarity 91.4%; Score 32; DB 1; Length 60;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7
DB 32 DSGELKR 38

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:15:35 ; Search time 69.1667 Seconds
(without alignments)
51.825 Million cell updates/sec

Title: US-10-646-919-37

Perfect score: 35
Sequence: 1 DSGELKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : uniprot_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	199	2	Q8PA66 xanthomonas
2	33	94.3	429	2	Q8EJS4 shewanella
3	33	94.3	643	2	Q9PP43 campylobact
4	32	91.4	60	1	RL32_BRAJA
5	32	91.4	60	1	RL32_RHOFA
6	32	91.4	62	1	RL32_METEX
7	32	91.4	330	2	Q96LH8
8	32	91.4	367	2	Q8NEB2
9	32	91.4	480	2	Q9LSK7
10	32	91.4	495	2	Q683K1 arabidopsi
11	32	91.4	495	2	Q8GW96 arabidopsi
12	32	91.4	617	2	Q6ZH27 burkholderi
13	32	91.4	617	2	Q6JQJ2 burkholderi
14	32	91.4	910	2	Q7RSD3 plasmodium
15	32	91.4	996	2	Q9S112 arabidopsi
16	31	88.6	216	2	Q8JDF7 coxiella bu
17	31	88.6	223	2	Q89GT6 bradyrhizob
18	31	88.6	248	1	Q8A522 bacteroides
19	31	88.6	384	1	YAGA_ECOLI
20	31	88.6	534	2	Q883N3 pseudomonas
21	31	88.6	612	2	Q9C927 arabidopsi
22	31	88.6	737	2	Q9CJS2 pasteurella
23	31	88.6	825	2	Q6FBW4 acinetobact
24	31	88.6	828	2	Q87KJ8 vibrio para
25	31	88.6	1028	1	CARB_THET2
26	31	88.6	1630	2	Q70073 thermus the
27	31	88.6	3940	2	Q9HEI9 anopheles g
28	31	88.6	3941	2	Q787K6 neurospora
29	30	85.7	75	2	Q70235 mus muscula
30	30	85.7	111	2	Q8M5F0 oryza sativ
31	30	85.7	111	2	Q7G656 oryza sativ

32	30	85.7	114	2	Q6UD64 uncultured
33	30	85.7	165	2	Q9AGB1 borrelia bu
34	30	85.7	172	2	Q8TLR8 methanobarc
35	30	85.7	172	2	Q8GR55 borrelia bu
36	30	85.7	185	2	Q31116 borrelia bu
37	30	85.7	188	2	Q6RU9 borrelia ga
38	30	85.7	188	2	Q6WQP5 borrelia bu
39	30	85.7	189	2	Q6WQP7 borrelia bu
40	30	85.7	191	2	Q31120 borrelia bu
41	30	85.7	192	2	Q9S3P2 borrelia bu
42	30	85.7	197	2	Q8GR57 borrelia va
43	30	85.7	200	2	Q86HP7 dictyostell
44	30	85.7	210	2	Q44719 borrelia bu
45	30	85.7	210	2	Q45176 borrelia ga

ALIGNMENTS

```

RESULT 1
ID Q8PA66 PRELIMINARY; PRT; 199 AA.
AC Q8PA66;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE Hypothetical protein XCC1623.
OS OrderedLocustNames=XCC1623;
GN Xanthomonas campestris (pv. Campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_Taxid=340;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33913 / NCPB 528;
RX MEDLINE=22022145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
RA Camarotte G., Canavan F., Cardozo J., Chamberg F., Ciapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Farra J.B., Ferreira A.U.S., Ferreira R.C.C., Ferro W.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidants J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira B.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463 (2002).
DR EMBL; AB01263; AAM40917.1; -.
KW Complete proteome.
SQ SEQUENCE 199 AA; 21899 MW; 6F9D797EC88B935B CRC64;

Query Match 100.0%; Score 35; DB 2; Length 199;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

THIS PAGE IS BLANK

Db 763 DSGELK 768

Search completed: July 20, 2005, 13:56:05
Job time : 15.5 secs

A/Map position: 1
C:Superfamily: Arabidopsis thaliana hypothetical protein At2g41080

Query Match
Best Local Similarity 85.7%; Score 30; DB 2; Length 665;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELK 7
|||||
DB 239 DSGELNR 245

RESULT 12
S48394
probable membrane protein Y1140w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C>Date: 02-Dec-1994 #sequence_rev1510n 02-Dec-1994 #text_change 09-Jul-2004
C:Accession: S48394; S50276
R:Churche, C.
submitted to the EMBL Data Library, September 1994
A:Reference number: S48310
A:Accession: S48394
A:Molecule type: DNA
A:Residues: 1-823 <CHU>
A:Cross-references: UNIPROT:P38928; GB:Z47047; EMBL:Z38059; NID:G603997; PID:G763206; MI
R:Torrey, L.E.; Gibbs, P.E.M.; Nelson, J.; Lawrence, C.W.
Yeast 10, 1503-1509, 1994
A>Title: Cloning and sequence of REV7, a gene whose function is required for DNA damage-
A:Reference number: S50275; MUID:95176709; PMID:7871890
A:Accession: S50276
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 80-823 <TOR>
A:Cross-references: EMBL:U07228; NID:G460247; PID:AAA67919.1; PID:G460249
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, March 1994
C:Genetics:
A:Gene: SGD:SRO4
A:Cross-references: SGD:S0001402; MIPS:Y1140w
A/Map position: 9L
C:Superfamily: Saccharomyces cerevisiae probable membrane protein Y1140w
C:Keywords: transmembrane protein
F:5-22/Domain: transmembrane #status predicted <TM1>
F:511-527/Domain: transmembrane #status predicted <TM2>

Query Match
Best Local Similarity 85.7%; Score 30; DB 2; Length 823;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6
|||||
DB 198 DSGELK 203

RESULT 13
A57238
iron-responsive element-binding protein 2, hepatic - rat
M:Alternate names: iron regulatory protein 2
C:Species: Rattus norvegicus (Norway rat)
C>Date: 08-Dec-1995 #sequence_rev1510n 08-Dec-1995 #text_change 16-Jul-1999
C:Accession: A57238
R:Guo, B.; Brown, F.M.; Phillips, J.D.; Yu, Y.; Leibold, E.A.
J. Biol. Chem. 270, 16529-16535, 1995
A>Title: Characterization and expression of iron regulatory protein 2 (IRP2). Presence of
A:Reference number: A57238; MUID:95348066; PMID:7622457
A:Accession: A57238
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-963 <GUO>
A:Cross-references: GB:U020181; NID:G897582; PID:AAA79927.1; PID:G897583
A:Experimental source: strain Sprague-Dawley; tissue type adult liver
C:Genetics:
A:Gene: IRP2
C:Superfamily: iron-responsive element-binding protein

C:Keywords: 4Fe-4S; iron-sulfur protein; liver; metalloprotein; RNA binding
F:555-698/Domain: RNA binding #status predicted <RNA>
F:512,578,581/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match
Best Local Similarity 85.7%; Score 30; DB 2; Length 963;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELK 7
|||||
DB 180 DSGELNR 186

RESULT 14
D97001
probable membrane protein [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_rev1510n 14-Sep-2001 #text_change 09-Jul-2004
C:Accession: D97001
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A>Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: D97001
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1116 <KOR>
A:Cross-references: UNIPROT:Q97KU3; GB:AE001437; PID:AAK78799.1; PID:G15023714; GSPDB:G
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0823

Query Match
Best Local Similarity 85.7%; Score 30; DB 2; Length 1116;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6
|||||
DB 865 DSGELK 870

RESULT 15
E86160
hypothetical protein F22D16.1 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_rev1510n 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: E86160
R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultz, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Frazer, C.W.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E86160
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1254 <STO>
A:Cross-references: UNIPROT:Q9SR21; GB:AE005172; NID:G6056403; PID:AAF02867.1; GSPDB:G
A/Map position: 1

Query Match
Best Local Similarity 85.7%; Score 30; DB 2; Length 1254;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6
|||||

A:Cross-references: UNIPROT:Q44976; GB:U25413; NID:g9495735; PIDN:AAA22956.1; PID:g9495736
C:Genetics:
A:Gene: ospC
C:Superfamily: Lyme disease epirochete surface protein C

Query Match 85.7%; Score 30; DB 2; Length 211;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DSGELKR 6
|||
Db 117 DSGELK 122

RESULT 8

B96759

protein serine carboxypeptidase T18K17.4 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 16-Aug-2004

C:Accession: B96759

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzeberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: B96759

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-438 <STO>

A:Cross-references: UNIPROT:Q9CAN2; GB:AE005173; NID:g6598854; PIDN:AAF18708.1; GSPDB:GN C:Genetics:

A:Gene: T18K17.4

A:Map position: 1

C:Superfamily: Serine carboxypeptidase

Query Match 85.7%; Score 30; DB 2; Length 438;
Best Local Similarity 85.7%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DSGELKR 7
|||
Db 151 DSGELK 157

RESULT 9

A96759

protein serine carboxypeptidase T18K17.5 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 16-Aug-2004

C:Accession: A96759

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzeberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: A96759

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-441 <STO>

A:Cross-references: UNIPROT:Q9CAN1; GB:AE005173; NID:g6598855; PIDN:AAF18709.1; GSPDB:GN C:Genetics:

A:Gene: T18K17.5
A:Map position: 1
C:Superfamily: serine carboxypeptidase

Query Match 85.7%; Score 30; DB 2; Length 441;
Best Local Similarity 85.7%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DSGELKR 7
|||
Db 152 DSGELK 158

RESULT 10

C96759

protein serine carboxypeptidase T18K17.3 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 16-Aug-2004

C:Accession: C96759

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzeberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: C96759

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-441 <STO>

A:Cross-references: UNIPROT:Q9CAN3; GB:AE005173; NID:g6598853; PIDN:AAF18707.1; GSPDB:GN C:Genetics:

A:Gene: T18K17.3

A:Map position: 1

C:Superfamily: Serine carboxypeptidase

Query Match 85.7%; Score 30; DB 2; Length 441;
Best Local Similarity 85.7%; Pred. No. 96;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DSGELKR 7
|||
Db 152 DSGELK 158

RESULT 11

F86181

protein F13M7.17 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C:Accession: F86181

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzeberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: F86181

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-665 <STO>

A:Cross-references: UNIPROT:Q9MAT2; GB:AE005172; NID:g7211995; PIDN:AAF40466.1; GSPDB:GN C:Genetics:

A:Gene: F13M7.17

Best Local Similarity 85.7%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 7
|||||

DB 281 DNGELKR 287

RESULT 3

C64752
outer surface protein C precursor - *Escherichia coli* (strain K-12)

C/Species: *Escherichia coli*

C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004

C/Accession: C64752

R/Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Ch

A.; Rose, D.J.; Mau, B.; Shaoh, Y.

Science 277, 1453-1462, 1997

A/Title: The complete genome sequence of *Escherichia coli* K-12.

A/Reference number: A64720; PMID:9742617; PMID:9278503

A/Accession: C64752

A/Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-384 <BLAT>

A/Cross-references: UNIPROT:P37007; GB:AE000134; GB:U00096; NID:g1786454; PIDN:AACT3370.

A/Experimental source: strain K-12, substrain MG1655

C/Genetics:

A/Gene: y8gA

Query Match 88.6%; Score 31; DB 2; Length 384;

Best Local Similarity 85.7%; Pred. No. 50;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 7
|||||

DB 266 DSGELQR 272

RESULT 4

I40274
outer surface protein C precursor - *Borrelia* sp.

C/Species: *Borrelia* sp.

C/Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 26-May-2000

C/Accession: I40274

R/Fukunaga, M.; Hamase, A.

J. Clin. Microbiol. 33, 2415-2420, 1995

A/Title: Outer surface protein C gene sequence analysis of *Borrelia burgdorferi* sensu la

A/Reference number: I40269; PMID:96025162; PMID:7494039

A/Accession: I40274

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-210 <RES>

A/Cross-references: GB:D49381; NID:g1041109; PIDN:BAA08379.1; PID:g1041110

C/Superfamily: Lyme disease spirochete surface protein C

Query Match 85.7%; Score 30; DB 2; Length 210;

Best Local Similarity 100.0%; Pred. No. 44;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6
|||||

DB 114 DSGELK 119

RESULT 5

I40284
outer surface protein C precursor - *Borrelia* sp.

C/Species: *Borrelia* sp.

C/Date: 04-Sep-1997 #sequence_revision 04-Sep-1997 #text_change 26-May-2000

C/Accession: I40284; I40283

R/Fukunaga, M.; Hamase, A.

J. Clin. Microbiol. 33, 2415-2420, 1995

A/Title: Outer surface protein C gene sequence analysis of *Borrelia burgdorferi* sensu la

A/Reference number: I40269; PMID:96025162; PMID:7494039

A/Accession: I40284

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-210 <RES>

A/Cross-references: GB:D49508; NID:g707103; PIDN:BAA08468.1; PID:g763695

A/Accession: I40283

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-210 <RE2>

A/Cross-references: GB:D49507; NID:g707102; PID:g763694

C/Superfamily: Lyme disease spirochete surface protein C

Query Match 85.7%; Score 30; DB 2; Length 210;

Best Local Similarity 100.0%; Pred. No. 44;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6
|||||

DB 114 DSGELK 119

RESULT 6

I40144
outer surface protein C precursor - Lyme disease spirochete

C/Species: *Borrelia burgdorferi* (Lyme disease spirochete)

C/Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 09-Jul-2004

C/Accession: I40144; S70282

R/Stevenson, B.; Barthold, S.W.

FEMS Microbiol. Lett. 124, 367-372, 1994

A/Title: Expression and sequence of outer surface protein C among North American isolate

A/Reference number: I40143; PMID:95154673; PMID:7851744

A/Accession: I40144

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-210 <RES>

A/Cross-references: UNIPROT:Q44719; EMBL:U04281; NID:g434663; PIDN:AACT3297.1; PID:g4346

R/Livey, I.; Gibbs, C.P.; Schuster, R.; Dörner, F.

Mol. Microbiol. 18, 257-269, 1995

A/Title: Evidence for lateral transfer and recombination in OspC variation in Lyme disea

A/Reference number: S70255; PMID:96296448; PMID:8709845

A/Accession: S70282

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 19-210 <LIV>

A/Cross-references: EMBL:I42893; NID:g858721; PIDN:AAB37001.1; PID:g1695218

A/Experimental source: strain 297

C/Genetics:

C/Superfamily: Lyme disease spirochete surface protein C

Query Match 85.7%; Score 30; DB 2; Length 210;

Best Local Similarity 100.0%; Pred. No. 44;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6
|||||

DB 116 DSGELK 121

RESULT 7

I40268
outer surface protein C precursor - Lyme disease spirochete

C/Species: *Borrelia burgdorferi* (Lyme disease spirochete)

C/Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 09-Jul-2004

C/Accession: I40268

R/Margolis, N.; Hogan, D.; Cieplak, W.

Gene 143, 105-110, 1994

A/Title: Homology between *Borrelia burgdorferi* OspC and members of the family of *Borrelli*

A/Reference number: I40268; PMID:94259285; PMID:8200524

A/Accession: I40268

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-211 <RES>

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:32:16 ; Search time 14.5 Seconds
(without alignments)
46.449 Million cell updates/sec

Title: US-10-646-919-37

Perfect score: 35
Sequence: 1 DSGELKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	94.3	643	2 H81361	ABC transport syst
2	32	91.4	996	2 D84561	probable AAA-type
3	31	88.6	384	2 C64752	Yaga protein - Bsc
4	30	85.7	210	2 I40274	outer surface prot
5	30	85.7	210	2 I40284	outer surface prot
6	30	85.7	210	2 I40144	outer surface prot
7	30	85.7	211	2 I40268	outer surface prot
8	30	85.7	438	2 B96759	protein serine car
9	30	85.7	441	2 A96759	protein serine car
10	30	85.7	441	2 C96759	protein serine car
11	30	85.7	665	2 F86181	protein serine car
12	30	85.7	823	2 S48394	protein F13M7.17 l
13	30	85.7	963	2 A57238	probable membrane
14	30	85.7	1116	2 D97001	iron-responsive el
15	30	85.7	1254	2 E86160	probable membrane
16	30	85.7	1250	2 UC5473	hypothetical prote
17	30	85.7	1613	2 T06678	dextranucrase (EC
18	30	85.7	1871	2 D96796	hypothetical prote
19	30	85.7	305	2 AD0123	probable heat choc
20	29	82.9	3705	2 AD0123	probable autotrans
21	29	82.9	60	2 E87462	ribosomal protein
22	29	82.9	217	2 B97948	hypothetical prote
23	29	82.9	235	2 A71236	probable biotin-la
24	29	82.9	349	2 S70229	TYA protein - yeas
25	29	82.9	357	2 F83195	probable methylate
26	29	82.9	379	2 T44615	citrate (si)-synth
27	29	82.9	405	2 B85361	hypothetical prote
28	29	82.9	430	2 B84619	probable serine ca
29	29	82.9	430	2 F81983	probable valine-py
29	29	82.9	430	2 H81039	valine-pyruvate am

30	29	82.9	433	2 A84619	probable serine ca
31	29	82.9	438	2 S70231	TYA protein - yeas
32	29	82.9	438	2 S70232	TYA protein - yeas
33	29	82.9	438	2 S19347	TYA protein - yeas
34	29	82.9	438	2 S61574	TYA protein - yeas
35	29	82.9	438	2 S45397	TYA protein - yeas
36	29	82.9	438	2 S67320	TYA protein - yeas
37	29	82.9	438	2 S69968	TYA protein - yeas
38	29	82.9	438	2 S69965	TYA protein - yeas
39	29	82.9	438	2 S61589	TYA protein - yeas
40	29	82.9	438	2 F69440	conserved hypothet
41	29	82.9	461	2 E75304	hypothetical prote
42	29	82.9	472	2 G83537	aromatic amino aci
43	29	82.9	478	2 S69974	TYA protein - yeas
44	29	82.9	490	2 S76169	hypothetical prote
45	29	82.9	497	2 G95398	probable ABC trans

ALIGNMENTS

RESULT 1
H81361
ABC transport system ATP-binding protein Cj0888c [imported] - Campylobacter jejuni (scr:
CjSpecies: Campylobacter jejuni
CjDate: 31-Mar-2000 #sequence_revision 31-Mar-2000 #ext_change 16-Aug-2004
CjAccession: H81361
RjParkhill, J.; Wren, B.W.; Mungall, K.; Kettle, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrell
Nature 403, 665-668, 2000
A>Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hy
A:Reference number: AB1250; MUID:20150912; PMID:10688204
A:Accession: H81361
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-643 <PAR>
A:Cross-references: UNIPROT:Q9PP43; GB:AL139076; GB:AL111166; NID:96968128; PIDN:CAB731-
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj0888c
C:Superfamily: ATP-binding cassette homology

Query Match 94.3%; Score 33; DB 2; Length 643;
Best Local Similarity 85.7%; Pred. No. 30;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7
DB 384 DSGELKR 390

RESULT 2
D84561
probable AAA-type ATPase [imported] - Arabidopsis thaliana
CjSpecies: Arabidopsis thaliana (mouse-ear cress)
CjDate: 02-Feb-2001 #sequence_revision 02-Feb-2001 #ext_change 09-Jul-2004
CjAccession: D84561
RjLin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentto, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umeyam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eilen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter,
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:10617197
A:Accession: D84561
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-996 <STO>
A:Cross-references: UNIPROT:Q9S112; GB:AE002093; NID:94874284; PIDN:AAD31347.1; GSPDB:G
C:Genetics:
A:Gene: At2g18190
A:Map position: 2

Query Match 91.4%; Score 32; DB 2; Length 996;

```
RESULT 2
US-10-282-122A-54450
; Sequence 54450, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54450
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
US-10-282-122A-54450

Query Match          94.3%; Score 33; DB 15; Length 643;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSGELKR 7
DB      384 DSGELKR 390

RESULT 3
US-09-764-891-3622
; Sequence 3622, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antipodides
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
```

```
; SEQ ID NO 3622
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (6)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (54)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (77)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-3622

Query Match          91.4%; Score 32; DB 10; Length 80;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSGELKR 7
DB      47 DSGELKR 53

RESULT 4
US-10-424-599-234622
; Sequence 234622, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 234622
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53892C.1.pep
US-10-424-599-234622

Query Match          91.4%; Score 32; DB 15; Length 287;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSGELKR 7
DB      122 DSGDLKR 128

RESULT 5
US-10-425-114-44704
; Sequence 44704, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven B
; APPLICANT: Tabaka, Jack B
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:52:31 ; Search time 71.3333 Seconds
(without alignments)
38.093 Million cell updates/sec

Title: US-10-646-919-37

Perfect score: 35
Sequence: 1 DSGELKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1736639 seqs, 388189149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubppaa/US11_PUBCOMB.pep:*
20: /cgn2_6/ptodata/2/pubppaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	7	US-10-646-919-37	Sequence 37, Appl
2	33	94.3	643	US-10-282-122A-54450	Sequence 54450, A
3	33	91.4	80	US-09-764-891-3622	Sequence 3622, Ap
4	32	91.4	287	US-10-424-599-234622	Sequence 234622, A
5	32	91.4	292	US-10-425-114-44704	Sequence 44704, A
6	32	91.4	339	US-10-188-248-82	Sequence 82, Appl
7	32	91.4	340	US-10-188-248-84	Sequence 84, Appl
8	32	91.4	480	US-10-225-066A-496	Sequence 496, App
9	32	91.4	480	US-10-374-780A-2902	Sequence 2902, Ap
10	32	91.4	480	US-10-669-824-60	Sequence 60, Appl
11	32	91.4	480	US-10-870-198-60	Sequence 60, Appl

12	32	91.4	522	15	US-10-425-114-56109	Sequence 56109, A
13	31	88.6	257	15	US-10-424-599-163059	Sequence 163059, A
14	30	85.7	99	16	US-10-767-701-31966	Sequence 31966, A
15	30	85.7	111	16	US-10-437-963-143516	Sequence 143516, A
16	30	85.7	120	16	US-10-437-963-145261	Sequence 145261, A
17	30	85.7	208	16	US-10-437-963-173203	Sequence 173203, A
18	30	85.7	222	16	US-10-437-963-113982	Sequence 113982, A
19	30	85.7	270	16	US-10-767-701-38282	Sequence 38282, A
20	30	85.7	283	15	US-10-369-493-13574	Sequence 13574, A
21	30	85.7	300	9	US-09-738-626-4711	Sequence 4711, Ap
22	30	85.7	307	16	US-10-425-115-244452	Sequence 244452, A
23	30	85.7	341	15	US-10-437-963-131112	Sequence 131112, A
24	30	85.7	415	15	US-10-369-493-7947	Sequence 7947, Ap
25	30	85.7	426	15	US-10-425-114-53923	Sequence 53923, A
26	30	85.7	463	17	US-10-470-048B-577	Sequence 577, App
27	30	85.7	465	15	US-10-282-122A-70724	Sequence 70724, A
28	30	85.7	465	18	US-10-724-972A-5599	Sequence 5599, Ap
29	30	85.7	499	16	US-10-767-701-46391	Sequence 46391, A
30	30	85.7	511	16	US-10-437-963-125840	Sequence 125840, A
31	30	85.7	526	15	US-10-425-114-62867	Sequence 62867, A
32	30	85.7	583	16	US-10-437-963-114326	Sequence 114326, A
33	30	85.7	584	16	US-10-425-115-348909	Sequence 348909, A
34	30	85.7	594	16	US-10-437-963-173742	Sequence 173742, A
35	30	85.7	779	16	US-10-437-963-115617	Sequence 115617, A
36	30	85.7	963	9	US-09-924-396B-20	Sequence 20, Appl
37	30	85.7	3705	15	US-10-282-122A-77944	Sequence 77944, A
38	29	82.9	7	18	US-10-646-919-53	Sequence 53, Appl
39	29	82.9	82	15	US-10-424-599-180374	Sequence 180374, A
40	29	82.9	103	16	US-10-425-115-258364	Sequence 258364, A
41	29	82.9	105	16	US-10-425-115-225746	Sequence 225746, A
42	29	82.9	106	17	US-10-732-923-1434	Sequence 12434, A
43	29	82.9	159	15	US-10-424-599-227781	Sequence 227781, A
44	29	82.9	170	15	US-10-424-599-171900	Sequence 171900, A
45	29	82.9	187	16	US-10-437-963-125509	Sequence 125509, A

ALIGNMENTS

RESULT 1
US-10-646-919-37
Sequence 37, Application US/10646919
Publication No. US20050148075A1
GENERAL INFORMATION:
APPLICANT: Barbas, Carlos F.
TITLE OF INVENTION: Zinc-finger binding domains for GNN
FILE REFERENCE: TSRI 645.2C1
CURRENT APPLICATION NUMBER: US/10/646,919
CURRENT FILING DATE: 2003-08-21
PRIOR APPLICATION NUMBER: US 09/494,190
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: PCT/EP99/07742
PRIOR FILING DATE: 1999-10-14
PRIOR APPLICATION NUMBER: US 09/173,941
PRIOR FILING DATE: 1998-10-16
NUMBER OF SEQ ID NOS: 129
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 37
LENGTH: 7
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthesized
US-10-646-919-37

Query Match 100.0%; Score 35; DB 18; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.6e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7
|||||
DB 1 DSGELKR 7

THIS PAGE IS BLANK

Qy	1	DSGEIKR	7
Db	200	DSGEIKR	206

Search completed: July 20, 2005, 13:54:28
Job time : 22.6667 secs


```

; LENGTH: 224
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (76)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-482-273-174

Query Match
Best Local Similarity 82.9%; Score 29; DB 4; Length 224;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELKR 7
DB 188 DLGELKR 194

RESULT 12
US-09-902-540-10834
; Sequence 10834, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 10834
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Myxococcus xanthus
US-09-902-540-10834

Query Match
Best Local Similarity 82.9%; Score 29; DB 4; Length 300;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7
DB 194 DAGELKR 200

RESULT 13
US-09-482-273-105
; Sequence 105, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; PRIOR FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 105
; LENGTH: 302
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```

US-09-482-273-105

Query Match
Best Local Similarity 82.9%; Score 29; DB 4; Length 302;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELKR 7
DB 188 DLGELKR 194

RESULT 14
US-09-252-991A-31133
; Sequence 31133, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31133
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31133

Query Match
Best Local Similarity 82.9%; Score 29; DB 4; Length 365;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SGEELKR 7
DB 169 SGEELKR 174

RESULT 15
US-09-538-092-21
; Sequence 21, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loti
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 21
; LENGTH: 438
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YBL101W-A
US-09-538-092-21

Query Match
Best Local Similarity 82.9%; Score 29; DB 4; Length 438;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
RESULT 7
; US-09-710-279-658
; Sequence 658, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 658
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-09-710-279-658
```

```
Query Match      85.7%; Score 30; DB 4; Length 472;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 DSGELK 6
        |||||
Db      401 DSGELK 406
```

```
RESULT 8
; US-09-173-941-53
; Sequence 53, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV0081S
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 53
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
US-09-173-941-53
```

```
Query Match      82.9%; Score 29; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      2 SGE LK R 7
        |||||
Db      2 SGE LK R 7
```

```
RESULT 9
; US-09-494-190-53
; Sequence 53, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
```

```
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1999-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 53
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:phage display
US-09-494-190-53
```

```
Query Match      82.9%; Score 29; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      2 SGE LK R 7
        |||||
Db      2 SGE LK R 7
```

```
RESULT 10
; US-09-270-767-60792
; Sequence 60792, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 60792
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Xaa means any amino acid
US-09-270-767-60792
```

```
Query Match      82.9%; Score 29; DB 4; Length 66;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      2 SGE LK R 7
        |||||
Db      60 SGE LK R 65
```

```
RESULT 11
; US-09-482-273-174
; Sequence 174, Application US/09482273
; Patent No. 6534631
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 71 Human Secreted Proteins
; FILE REFERENCE: P2030P1
; CURRENT APPLICATION NUMBER: US/09/482,273
; CURRENT FILING DATE: 2000-01-13
; EARLIER APPLICATION NUMBER: PCT/US99/15849
; EARLIER FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,922
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/092,956
; EARLIER FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 267
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 174
```

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:phage display
OTHER INFORMATION: selected and mutagenized
US-09-494-190-37

Query Match 100.0%; Score 35; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 7
|||||
DB 1 DSGELK 7

RESULT 3
US-09-248-796A-18533
Sequence 18533, Application US/09248796A

PATENT No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A

PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725

PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409

NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 18533
LENGTH: 612
TYPE: PRT
ORGANISM: Candida albicans

US-09-248-796A-18533
Query Match 88.6%; Score 31; DB 4; Length 612;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELK 7
|||||
DB 580 DSGELK 586

RESULT 4
US-08-671-548C-12
Sequence 12, Application US/08671548C

PATENT No. 6486130
GENERAL INFORMATION:

APPLICANT: LIVEY, Ian
APPLICANT: CROWE, Brian
TITLE OF INVENTION: IMMUNOGENIC FORMULATION OF OSCP ANTIGEN VACCINE FOR THE PREVENTIO
TITLE OF INVENTION: TREATMENT OF LYME DISEASE AND RECOMBINANT METHODS FOR THE PREPAR
FILE REFERENCE: 37974-0023
CURRENT APPLICATION NUMBER: US/08/671,548C

PRIOR FILING DATE: 2001-08-30
PRIOR APPLICATION NUMBER: US 08/284,667

PRIOR FILING DATE: 1994-08-19
PRIOR APPLICATION NUMBER: 08/053,863

PRIOR FILING DATE: 1993-04-29
PRIOR APPLICATION NUMBER: PCT/EP94/01365

NUMBER OF SEQ ID NOS: 78
SOFTWARE: Patentin version 3.0
SEQ ID NO 12
LENGTH: 192
TYPE: PRT
ORGANISM: Borrelia burgdorferi
US-08-671-548C-12

Query Match 85.7%; Score 30; DB 4; Length 192;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6
|||||
DB 98 DSGELK 103

RESULT 5
US-09-345-236B-43
Sequence 43, Application US/09345236B

PATENT No. 6521454
GENERAL INFORMATION:
APPLICANT: Becnel, James J.
APPLICANT: Tokuo, Fukuda

APPLICANT: Moser, Bettina
APPLICANT: Cockburn, Andrew

APPLICANT: White, Susan E.
APPLICANT: Undeen, Albert H.
TITLE OF INVENTION: No. 6521454e1 Baculoviruses, Insecticidal
FILE REFERENCE: 21042,0004
CURRENT APPLICATION NUMBER: US/09/345,236B

PRIOR FILING DATE: 1999-06-30
NUMBER OF SEQ ID NOS: 148
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 43
LENGTH: 433
TYPE: PRT
ORGANISM: mosquito baculovirus

US-09-345-236B-43
Query Match 85.7%; Score 30; DB 4; Length 433;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELK 7
|||||
DB 130 DSGELK 136

RESULT 6
US-09-134-001C-3856
Sequence 3856, Application US/09134001C

PATENT No. 6380370
GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C

PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3856
LENGTH: 465
TYPE: PRT
ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3856
Query Match 85.7%; Score 30; DB 3; Length 465;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELK 6
|||||
DB 401 DSGELK 406

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:17:05 ; Search time 21.6667 Seconds
(without alignments)
24.117 Million cell updates/sec

Title: US-10-646-919-37

Perfect score: 35

Sequence: 1 DSGELKR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	7	US-09-173-941-37	Sequence 37, App1
2	35	100.0	7	US-09-494-190-37	Sequence 37, App1
3	31	88.6	612	US-09-248-796A-18533	Sequence 18533, A
4	30	85.7	192	US-08-671-548C-12	Sequence 12, App1
5	30	85.7	433	US-09-345-236B-43	Sequence 43, App1
6	30	85.7	465	US-09-134-001C-3896	Sequence 3856, Ap
7	30	85.7	472	US-09-710-279-658	Sequence 658, App
8	29	82.9	7	US-09-173-941-53	Sequence 53, App1
9	29	82.9	7	US-09-494-190-53	Sequence 53, App1
10	29	82.9	66	US-09-270-767-60792	Sequence 60792, A
11	29	82.9	224	US-09-482-273-174	Sequence 174, App
12	29	82.9	300	US-09-902-540-10834	Sequence 10834, A
13	29	82.9	302	US-09-482-273-105	Sequence 105, App
14	29	82.9	365	US-09-252-991A-31133	Sequence 31133, A
15	29	82.9	438	US-09-538-092-21	Sequence 21, App1
16	29	82.9	438	US-09-538-092-62	Sequence 62, App1
17	29	82.9	475	US-09-252-991A-29697	Sequence 29697, A
18	29	82.9	640	US-09-907-794A-292	Sequence 292, App
19	29	82.9	640	US-09-905-125A-292	Sequence 292, App
20	29	82.9	640	US-09-902-775A-292	Sequence 292, App
21	29	82.9	640	US-09-906-700-292	Sequence 292, App
22	29	82.9	640	US-09-903-603A-292	Sequence 292, App
23	29	82.9	640	US-09-904-920A-292	Sequence 292, App
24	29	82.9	640	US-09-909-064-292	Sequence 292, App
25	29	82.9	640	US-09-905-381A-292	Sequence 292, App
26	29	82.9	640	US-09-906-618-292	Sequence 292, App
27	29	82.9	718	US-09-252-991A-25696	Sequence 25696, A

28	29	82.9	830	4	US-09-252-991A-22004	Sequence 22004, A
29	29	82.9	1191	4	US-09-921-099A-2	Sequence 2, App1
30	29	82.9	1191	4	US-09-921-099A-4	Sequence 36, App1
31	28	80.0	7	3	US-09-173-941-36	Sequence 36, App1
32	28	80.0	7	4	US-09-494-190-36	Sequence 64, App1
33	28	80.0	43	2	US-08-488-161-64	Sequence 64, App1
34	28	80.0	43	3	US-09-273-685-64	Sequence 64, App1
35	28	80.0	43	5	PCT-US95-11934-64	Sequence 64, App1
36	28	80.0	139	4	US-09-402-016A-2	Sequence 8288, Ap
37	28	80.0	181	4	US-09-489-039A-8288	Sequence 8288, Ap
38	28	80.0	201	4	US-09-786-256C-5	Sequence 5, App1
39	28	80.0	206	4	US-09-270-767-36554	Sequence 36554, A
40	28	80.0	206	4	US-09-270-767-51771	Sequence 51771, A
41	28	80.0	311	4	US-09-902-540-11886	Sequence 11886, A
42	28	80.0	332	4	US-09-902-540-11886	Sequence 9978, Ap
43	28	80.0	358	4	US-09-784-508-4	Sequence 4, App1
44	28	80.0	436	3	US-08-669-378-8	Sequence 8, App1
45	28	80.0	465	4	US-09-252-991A-21223	Sequence 21223, A

ALIGNMENTS

```
RESULT 1
US-09-173-941-37
; Sequence 37, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV00815
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
US-09-173-941-37

Query Match          100.0%; Score 35; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 4; 1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSGELKR 7
      |||||
Db      1 DSGELKR 7

RESULT 2
US-09-494-190-37
; Sequence 37, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231342P.
 PR 08-SEP-2000; 2000US-0231343P.
 PR 08-SEP-2000; 2000US-0231444P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234597P.
 PR 25-SEP-2000; 2000US-0234598P.
 PR 26-SEP-2000; 2000US-0234584P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235835P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239335P.
 PR 13-OCT-2000; 2000US-0239337P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-02446475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.

PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX WPI; 2001-465570/50.
 DR N-PSDB; AAL00934.
 XX
 PT Isolated nucleic acid molecule encoding a reproductive system antigen is
 XX used in preventing, treating or ameliorating a medical condition.
 PS Claim 11; SEQ ID NO 3622; 1297bp + Sequence Listing; English.
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a protein of the invention
 XX
 SQ Sequence 80 AA;
 QY 1 DSGELKR 7
 Db 47 DSGELKR 53
 QY 1 DSGELKR 7
 Db 47 DSGELKR 53
 RESULT 6
 ABB95674
 ID ABB95674 standard; protein; 80 AA.
 XX
 AC ABB95674;
 XX
 DT 21-JUN-2002 (first entry)
 DE Human testicular antigen SEQ ID NO: 1058.
 XX
 KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
 KW reproductive system disorder; urinary system disorder; gene therapy;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 XX gastrointestinal disease; infection; cystostatic.
 OS Homo sapiens.
 XX
 PN MO200155317-A2.
 XX
 PD 02-AUG-2001.
 XX

RESULT 4
ABU26526
ID ABU26526 standard; protein; 643 AA.
XX
AC ABU26526;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by prokaryotic essential gene #12053.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Campylobacter jejuni.
XX
PN WO20027183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind WJ,
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR MPI: 2003-0299926/02.
DR N-PSDB; ACA30396.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 54450; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 643 AA;
XX
Query Match 94.3%; Score 33; DB 6; Length 643;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 DSGELKR 7
DB 384 DSGELKR 390
XX
RESULT 5
AAM94964
ID AAM94964 standard; protein; 80 AA.
XX
AC AAM94964;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen SEQ ID NO: 3622.
XX
DE Human reproductive system related antigen; reproductive system disorder;
XX
KW Human; reproductive system related antigen; reproductive system disorder;
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PI 17-JAN-2001; 2001WO-US001339.
XX
PF 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-SEP-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.

CC achieved through specific contacts from side chains of amino acid
CC residues in the alpha helix. Each zinc finger can recognise a sub-site of
CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains
CC allows the recognition of extended contiguous asymmetric DNA sequences.
CC For example, a synthetic polynucleotide protein containing six zinc finger
CC domains can recognise an 18 bp sequence, and such proteins are
CC potentially highly gene-specific. The novel nucleotide-binding zinc
CC finger proteins may therefore be used in the development of artificial
CC gene-specific transcriptional regulators. Such transcriptional switches
CC may be used to regulate the expression of oncogenes such as erbB-2,
CC overexpression of which is involved in malignant transformation. The
CC proteins are therefore useful in the treatment of cancers, and may also
CC be used to activate genes involved in fighting diseases, and to treat
CC viral infections by inhibiting the synthesis of viral gene products. They
CC may be used in DNA-based diagnostic applications. The proteins may also
CC be used in producing functional gene knockout or activation in
CC heterozygous transgenic animals. Proteins of the invention can
CC discriminate between sequences which have a single base difference.
CC This is manifested in a >100-fold decrease in affinity for the variant
CC sequence. Gene activation and repression can be achieved by targeting
CC within the gene transcript, suggesting that information obtained from
CC expressed sequence tags may be sufficient for the construction of gene
CC switches. Sequences AAB02876-B02869 represent zinc finger alpha helix
CC phage library peptides disclosed in the invention
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 35; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7
| | | | |
DB 1 DSGELKR 7

RESULT 2

ADP61986
ID ADP61986 standard; peptide; 7 AA.

XX ADF61986;

XX 12-FEB-2004 (first entry)

XX Zinc finger binding region #37.

XX zinc finger-nucleotide binding polypeptide; expression regulation;

XX zinc finger binding region.

XX Synthetic.

XX US6610512-B1.

XX 26-AUG-2003.

XX 28-JUN-2000; 2000US-00494190.

XX 16-OCT-1998; 98US-00173941.

XX 14-OCT-1999; 99WO-EP007742.

XX (SCRI) SCRIPPS RES INST.

XX Barbas CF;

XX WPI; 2003-800134/75.

XX Regulating expression of nucleotide sequence that contains sequence 5'-
XX (GNN)n-3', comprises exposing nucleotide sequence to composition
XX comprising isolated and purified zinc finger-nucleotide binding
XX polypeptide.

XX Disclosure; SEQ ID NO 37; 46pp; English.

CC The invention relates to an expression of nucleotide sequence that
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by
CC exposing nucleotide sequence to composition comprising isolated and
CC purified zinc finger-nucleotide binding polypeptide that contains a
CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Ileu-Val-
CC Arg. The invention is used in the regulation of the expression of the
CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present
CC sequence represents the amino acid sequence of a zinc finger binding
CC region.
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 35; DB 7; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7
| | | | |
DB 1 DSGELKR 7

RESULT 3

ADJ98428
ID ADJ98428 standard; peptide; 7 AA.

XX ADJ98428;

XX 06-MAY-2004 (first entry)

XX Zinc finger DNA binding peptide #99.

XX library; multimeric DNA binding polypeptide;

XX zinc finger DNA binding peptide; gene expression silencing;

XX gene expression enhancement.

XX Unidentified.

XX WO2003066828-A2.

XX 14-AUG-2003.

XX 07-FEB-2003; 2003WO-US003705.

XX 07-FEB-2002; 2002US-0354981P.

XX (SCRI) SCRIPPS RES INST.

XX Barbas CF, Blancafort P;

XX WPI; 2003-731499/69.

XX New zinc finger library of multimeric DNA binding polypeptides, useful
XX for sterically occluding the binding site of a natural transcription
XX factor, and enhancing or silencing target gene expression.

XX Disclosure; Fig 20; 64pp; English.

XX The invention comprises a library of multimeric DNA binding polypeptides
XX (e.g. zinc finger DNA binding polypeptides). The zinc finger binding
XX polypeptides of the invention are useful for sterically occluding the
XX binding site of a natural transcription factor and enhancing or silencing
XX target gene expression. The present amino acid sequence represents a zinc
XX finger DNA binding peptide of the invention.

SQ Sequence 7 AA;

Query Match 100.0%; Score 35; DB 7; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELKR 7
| | | | |
DB 1 DSGELKR 7

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:14:30 ; Search time 82.5 Seconds
(without alignments)
32.816 Million cell updates/sec

Title: US-10-646-919-37

Perfect score: 35
Sequence: 1 DSGELKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_16Dec04:*
1: geneseqp19808:*
2: geneseqp19808:*
3: geneseqp20008:*
4: geneseqp20018:*
5: geneseqp20028:*
6: geneseqp20038:*
7: geneseqp20038:*
8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	35	100.0	7	3	AAB02896	Aab02896 Nucleotid
2	35	100.0	7	7	ADP61986	Adp61986 zinc fing
3	35	100.0	7	7	ADJ98428	Adj98428 zinc fing
4	33	94.3	643	6	ABU26526	Abu26526 Protein e
5	32	91.4	80	4	AAM94964	Aam94964 Human rep
6	32	91.4	80	4	ABB95674	Abb95674 Human tes
7	32	91.4	104	4	AAU65048	Aau65048 Propionib
8	32	91.4	104	6	ABM61567	Abm61567 Propionib
9	32	91.4	339	7	ADC13603	Adc13603 Human NOV
10	32	91.4	340	7	ADC13605	Adc13605 Human NOV
11	32	91.4	405	3	AAG13174	Aag13174 Arabidops
12	32	91.4	410	8	ADN73063	Adn73063 Thale cre
13	32	91.4	413	3	AAG13173	Aag13173 Arabidops
14	32	91.4	422	3	AAG13172	Aag13172 Arabidops
15	32	91.4	480	7	ADD30464	Add30464 Plant yie
16	32	91.4	480	8	AD144439	Adi14439 Plant tra
17	32	91.4	480	8	ADQ16268	Adq16268 Thalecres
18	32	91.4	966	5	ABB91788	Abb91788 Herbicida
19	31	88.6	992	3	AAG38672	Aag38672 Arabidops
20	31	88.6	1009	3	AAG38671	Aag38671 Arabidops
21	31	88.6	1027	3	AAG38670	Aag38670 Arabidops
22	30	85.7	99	4	AAU46854	Aau46854 Propionib
23	30	85.7	99	6	ABM43373	Abm43373 Propionib
24	30	85.7	120	4	AAU45935	Aau45935 Propionib
25	30	85.7	120	6	ABM42454	Abm42454 Propionib

26	30	85.7	176	2	AAB62775	Aab62775 Borrelia
27	30	85.7	190	4	AAB62707	Aab62707 B burgdor
28	30	85.7	192	2	AAB60889	Aab60889 Borrelia
29	30	85.7	192	6	ABU61170	Abu61170 Borrelia
30	30	85.7	206	6	ADA14376	Ada14376 Mouse spe
31	30	85.7	283	8	ADS24541	Ads24541 Bacterial
32	30	85.7	300	4	AAB79948	Aab79948 Coryneb
33	30	85.7	300	4	AAB79901	Aab79901 Coryneb
34	30	85.7	300	4	AAG90957	Aag90957 C glutam
35	30	85.7	363	7	ADM26713	Adm26713 Hyperther
36	30	85.7	368	4	AAB62718	Aab62718 Borrelia
37	30	85.7	370	4	AAB62715	Aab62715 Borrelia
38	30	85.7	374	4	AAB62710	Aab62710 Borrelia
39	30	85.7	378	4	AAB62712	Aab62712 Borrelia
40	30	85.7	392	4	AAB62734	Aab62734 Borrelia
41	30	85.7	394	4	AAB62730	Aab62730 Borrelia
42	30	85.7	398	4	AAB62728	Aab62728 Borrelia
43	30	85.7	401	4	AAB62738	Aab62738 Borrelia
44	30	85.7	415	8	ADN25294	Adn25294 Bacterial
45	30	85.7	433	7	ABO22403	AbO22403 Mosquito

ALIGNMENTS

RESULT 1	
AAB02896	
ID	AAB02896 standard; peptide; 7 AA.
XX	
AC	AAB02896;
DT	
18-SEP-2000	(first entry)
XX	
DE	Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:37.
XX	
KW	zinc finger domain; alpha helix; nucleotide binding; DNA binding;
KW	polydactyl protein; asymmetric target recognition;
KW	gene specific transcriptional regulator; gene activator; gene repressor;
KW	transcriptional switch; oncogene; erbB-2; cancer; tumour; gene therapy;
KW	transgenic animal; antiviral; anticancer; diagnosis.
XX	
OS	Synthetic.
XX	
PN	W0200023464-A2.
XX	
PD	27-APR-2000.
XX	
PF	14-OCT-1999; 99WO-EP007742.
XX	
PR	16-OCT-1998; 98US-00173941.
XX	
PA	(NOVS) NOVARTIS AG.
PA	(NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PA	(SCRT) SCRIPPS RES INST.
XX	
PI	Barbas CF;
XX	
DR	WPI; 2000-339648/29.
XX	
PT	Novel isolated and purified zinc finger nucleotide-binding proteins with
PT	specificity for GNN triplet sequences, useful in gene therapy and for
PT	regulating gene function.
XX	
PS	Disclosure; Fig 1; 48pp; English.
XX	
CC	The invention relates to zinc finger nucleotide-binding proteins which
CC	comprise 2-12, preferably 2-6, operatively linked motifs selected from
CC	sequences AAB02860-B02875. Sequences AAB02860-B02875 represent the alpha
CC	helical regions of zinc finger domains which specifically bind to target
CC	nucleotide triplets of the sequence 5'-GNN-3'. Such regions may be linked
CC	by the peptide linker TGEKP (AAB02970). The Cys2-His2 zinc finger motif
CC	is the most frequently utilised nucleic acid binding motif in eukaryotes,
CC	and constitutes a beta-beta-alpha fold. Nucleic acid recognition is


```
DT 01-MAR-2003 (TReMBLrel. 23, last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, last annotation update)
DE Hypothetical protein SO1800.
GN OrderedLocustNames=SO1800;
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=22297686; PubMed=12368613; DOI=10.1038/nbt749;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eilen J.A., Seshadri R., Ward N.L., Methe B.A.,
RA Bryant R.A., Meyer T., Tsapin A., Scott J., Beaman M.J.,
RA Harte D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A.,
RA White O., Wolf A.M., Vamathevan J.J., Weidman J.F., Impraim M.,
RA Lee K., Berry K.J., Lee C., Mueller J., Khoult H.M., Gill J.,
RA Uterback T.R., McDonald L.A., Feldblyum T.V., Smith H.O.,
RA Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL: AE015624; AAN54852.1; -.
DR TIGR: SO1800; -.
DR InterPro: IPR010758; SC_ADH.
DR Pfam: PF07055; scADH; 1.
DR Complete proteome.
SQ SEQUENCE 400 AA; 43912 MW; BE230180DA2D0868 CRC64;

Query Match
Best Local Similarity 100.0%; Score 34; DB 2; Length 400;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 148 DSGELVR 154

RESULT 3
ID Q70453 PRELIMINARY; PRT; 710 AA.
AC Q70453;
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, last annotation update)
DT 01-MAR-2004 (TReMBLrel. 26, last annotation update)
DE AGCP11556.
GN Name=agCG49309; ORFName=ENGANG0000019412;
OS Anopheles gambiae str. PE8T.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PE8T;
RA Anopheles Genome Sequencing Consortium;
RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RL -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AAB01008964; EAA12458.1; -.
DR HSP: P35222; IUDH.
DR InterPro: IPR008938; ARM.
DR InterPro: IPR000225; Armadillo.
DR InterPro: IPR000357; HEAT.
DR Pfam: PF00514; Arm; 6.
DR Pfam: PF02985; HEAT; 2.
DR PROSITE: PS50176; ARM_REPEAT; 1.
DR PROSITE: PS50176; ARM_REPEAT; 1.
SQ SEQUENCE 710 AA; 77053 MW; B934CD21FA02439 CRC64;

Query Match
Best Local Similarity 100.0%; Score 34; DB 2; Length 710;
Matches 100.0%; Pred. No. 1.1e+02;
```

```
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 535 DSGELVR 541

RESULT 4
ID Q8K6U1 PRELIMINARY; PRT; 448 AA.
AC Q8K6U1; Q79XM6;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, last annotation update)
DE Hypothetical phage associated protein SPYM3_1130 (Hypothetical protein
DE SP80734).
GN OrderedLocustNames=SP80734, SPYM3_1130;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus
OX NCBI_TaxID=198466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RX MEDLINE=22133808; PubMed=12122206; DOI=10.1073/pnas.152298499;
RA Bers S.B., Sylva G.L., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu W.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;
RT "Genome sequence of a serotype M3 strain of group A Streptococcus:
RT phage-encoded toxins, the high-virulence phenotype, and clone
RT emergence."
RT Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-1;
RX MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okanashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution."
RL Genome Res. 13:1042-1055(2003).
DR EMBL: AE014158; AAM79737.1; -.
DR EMBL: AP005143; BAC63829.1; -.
DR Complete proteome; Hypothetical protein.
SQ SEQUENCE 448 AA; 53074 MW; 9D1F78970DED7A05 CRC64;

Query Match
Best Local Similarity 97.1%; Score 33; DB 2; Length 448;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 361 DSGELVR 367

RESULT 5
ID Q726S1 PRELIMINARY; PRT; 550 AA.
AC Q726S1;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, last annotation update)
DE Methyl-accepting chemotaxis protein, putative.
GN OrderedLocustNames=DVU3035;
OS Desulfovibrio vulgaris (strain Hildenborough / ATCC 29579 / NCIMB
OS 8303).
OC Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales;
OC Desulfovibrionaceae; Desulfovibrio.
OX NCBI_TaxID=882;
RN [1]
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:15:35 ; Search time 69.1667 Seconds

(without alignments)
51.825 Million cell updates/sec

Title: US-10-646-919-36

Perfect score: 34

Sequence: 1 DSGELVR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_03.*

1: uniprot_sprot.*

2: uniprot_trcbl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	258	2 Q8BJP8	Q8BJP8 pseudomonas
2	34	100.0	400	2 Q8EG14	Q8EG14 shevanelia
3	34	100.0	710	2 Q7Q453	Q7Q453 anopheles g
4	33	97.1	448	2 Q8K6U1	Q8K6U1 streptococc
5	33	97.1	550	2 Q726S1	Q726S1 desulfovibr
6	31	91.2	91	2 P721S2	P721S2 pseudomonas
7	31	91.2	123	2 Q914N7	Q914N7 pseudomonas
8	31	91.2	132	2 Q61808	Q61808 pseudomonas
9	31	91.2	132	2 Q76EA7	Q76EA7 pseudomonas
10	31	91.2	132	2 Q884Y6	Q884Y6 pseudomonas
11	31	91.2	168	1 G8EB_XANCP	G8EB_XANCP xanthomonas
12	31	91.2	216	1 Q868N1	Q868N1 ixodes scap
13	31	91.2	299	2 Q9KRA4	Q9KRA4 vibrio chol
14	31	91.2	300	2 Q22090	Q22090 petunia hyl
15	31	91.2	315	2 Q63J16	Q63J16 burkholderi
16	31	91.2	355	2 Q8H5P5	Q8H5P5 oryza sativ
17	31	91.2	386	2 Q6D1E1	Q6D1E1 erythrina car
18	31	91.2	466	1 Q6MP11	Q6MP11 bdellovibr
19	31	91.2	477	1 MURC_XANAC	MURC_XANAC xanthomonas
20	31	91.2	477	1 MURC_XANCP	MURC_XANCP xanthomonas
21	31	91.2	650	2 Q62J76	Q62J76 burkholderi
22	31	91.2	650	2 Q63SW8	Q63SW8 burkholderi
23	31	91.2	685	2 Q9S8S0	Q9S8S0 arabidopsis
24	31	91.2	1079	2 Q758C5	Q758C5 neurospora
25	31	91.2	1286	2 Q752J0	Q752J0 ashyia gos
26	31	91.2	1290	2 Q48756	Q48756 leucosotoc
27	31	91.2	1952	2 Q955N5	Q955N5 drosophila
28	31	91.2	4547	2 Q9W343	Q9W343 drosophila
29	30	88.2	232	2 Q9BP45	Q9BP45 rhizobium 1
30	30	88.2	235	2 Q14195	Q14195 schizosacch
31	30	88.2	266	2 Q88F53	Q88F53 pseudomonas

32	30	88.2	289	2 Q983X2	Q983X2 rhizobium 1
33	30	88.2	290	2 Q89P18	Q89P18 bradyrhizob
34	30	88.2	292	2 Q6LNV2	Q6LNV2 photobacter
35	30	88.2	326	2 Q62141	Q62141 burkholderi
36	30	88.2	326	2 Q63RD9	Q63RD9 burkholderi
37	30	88.2	351	2 Q7NV61	Q7NV61 chromobacte
38	30	88.2	367	2 Q97212	Q97212 sulfobobus
39	30	88.2	447	2 Q15922	Q15922 leishmania
40	30	88.2	543	2 Q97U16	Q97U16 sulfolobus
41	30	88.2	642	2 Q72ER9	Q72ER9 desulfovibr
42	30	88.2	667	2 Q8KW91	Q8KW91 ruegeria sp
43	30	88.2	692	2 DNK1_SYNY3	DNK1_SYNY3
44	30	88.2	861	2 Q88N01	Q88N01 pseudomonas
45	30	88.2	873	1 MTS_CHRVO	MTS_CHRVO

ALIGNMENTS

RESULT 1					
Q8BJP8		PRELIMINARY;	PRT;	258 AA.	
AC	Q8BJP8:				
DT	01-JUN-2003 (TrEMBLrel. 24, Created)				
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)				
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)				
DE	Transcriptional regulator, IclR family.				
GN	OrderedLocNames=PP2601;				
OS	Pseudomonas putida (strain KT2440).				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
OC	Pseudomonadaceae; Pseudomonas.				
OX	NCBI_TaxID=160488;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=24243060; PubMed=12534463;				
RA	Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,				
RA	Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,				
RA	Brinkac L.M., Beaman M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,				
RA	Madupu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,				
RA	Hance I., Chris Lee P., Holtzapple B.K., Scanlan D., Tran K.,				
RA	Morzez A., Utechtack T.R., Rizzo M., Lee K., Kosack D., Moestl D.,				
RA	Wedler H., Lauber U., Stjepandic D., Hohnel J., Straetz M., Heim S.,				
RA	Kiewitz C., Eisen J.A., Timmis K.N., Duestehof A., Tummeler B.,				
RA	Fraser C.M.;				
RT	"Complete genome sequence and comparative analysis of the				
RL	metabolically versatile Pseudomonas putida KT2440."				
DR	Environ. Microbiol. 4:799-808(2002).				
DR	EMBL; AE016783; AA068209.1; -.				
DR	TIGR; PP2601; -.				
DR	GO; GO:0003677; F:DNA binding; IEA.				
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.				
DR	InterPro; IPR005471; HTH_ICLR.				
DR	InterPro; IPR009058; Wing_hlx_DNA_bnd.				
DR	PFam; PF01614; ICLR; 1.				
DR	SMART; SM00346; HTH_ICLR; 1.				
KW	Complete proteome; DNA-binding; Transcription;				
KW	Transcription regulation.				
SQ	SEQUENCE 258 AA; 27864 MW; 5411DFDCDE1772F CRC64;				
Query Match		100.0%;	Score 34;	DB 2;	Length 258;
Best Local Similarity		100.0%;	Pred. No. 39;		
Matches	7;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	1 DSGELVR 7				
DB	93 DSGELVR 99				
RESULT 2					
ID	Q8EG14	PRELIMINARY;	PRT;	400 AA.	
AC	Q8EG14;				
DT	01-MAR-2003 (TrEMBLrel. 23, Created)				

Query Match 91.2%; Score 31; DB 2; Length 299;
 Best Local Similarity 85.7%; Pred. No. 27;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
 |:|||||
 Db 46 DTGELVR 52

RESULT 3
 zinc finger protein ZPN3-3 [imported] - garden petunia
 C:Species: Petunia x hybrida (garden petunia)
 C>Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 09-Jul-2004
 C:Accession: T52379
 R:Kudo, K.; Sakamoto, A.; Kobayashi, A.; Rybka, Z.; Kanno, Y.; Nakagawa, H.; Mishino, T.
 Nucleic Acids Res. 26, 608-615, 1998
 A>Title: Cys2/His2 zinc-finger protein family of petunia: evolution and general mechanisms
 A:Reference number: Z26061
 A:Accession: T52379
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-300 <KUB>
 A:Cross-references: UNIPROT:Q22090; EMBL:AB006605; PDB:BAA21927.1
 A:Experimental source: strain Mitchell diploid
 C:Superfamily: Arabidopsis thaliana hypothetical protein F12B4.290

Query Match 91.2%; Score 31; DB 2; Length 300;
 Best Local Similarity 85.7%; Pred. No. 28;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
 |:|||||
 Db 154 DSGELVR 160

RESULT 4
 dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides
 C:Species: Leuconostoc mesenteroides
 C>Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 09-Jul-2004
 C:Accession: JCS473
 R:Monchois, V.; Willemot, R.M.; Remaud-Simeon, M.; Croux, C.; Monsan, P.
 Gene 182, 23-32, 1996
 A>Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconostoc
 A:Reference number: JCS473; MUID:97136686; PMID:8982063
 A:Accession: JCS473
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1290 <MON>
 A:Cross-references: UNIPROT:Q48756; GB:U38181
 C:Comment: This enzyme catalyses the transfer of D-glucopyranosyl units from sucrose onto
 C:Genetics:
 A:Gene: data
 C:Keywords: glycosyltransferase; hexosyltransferase
 F:8-870/Domain: catalytic #status predicted <CAT>
 F:922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 91.2%; Score 31; DB 2; Length 1290;
 Best Local Similarity 85.7%; Pred. No. 1.4e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
 |:|||||
 Db 1217 DTGELVR 1223

RESULT 5
 T38909
 hypothetical protein SPAC56E4.07 - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C:Accession: T38909
 R:Conor, R.; Churcher, C.M.; Barrrell, B.G.; Rajandream, M.A.; Wood, V.
 Submitted to the EMBL Data Library, September 1997
 A:Reference number: Z21813
 A:Accession: T38909
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-235 <CON>
 A:Cross-references: UNIPROT:Q14195; EMBL:Z99261; PDB:CAB1398.1; GSPDB:GN00066; SPDB:SF
 A:Experimental source: strain 97Zh-, cosmid G56E4
 C:Genetics:
 A:Gene: SPDB:SPAC56E4.07
 A:Map position: 1
 C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPAC56E4.07

Query Match 88.2%; Score 30; DB 2; Length 235;
 Best Local Similarity 71.4%; Pred. No. 36;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
 |:|||||
 Db 212 DTGELVR 218

RESULT 6
 C90506
 hypothetical protein ilvB-6 [imported] - Sulfolobus solfataricus
 C:Species: Sulfolobus solfataricus
 C>Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
 C:Accession: C90506
 R:She, Q.; Singh, R.K.; Confolenter, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-
 arett, R.A.; Kagan, M.A.; Jensen, C.W.; Van der Oost, J.
 J. J. Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
 Submitted to Genbank, April 2001
 A:Description: Sulfolobus solfataricus complete genome.
 A:Reference number: A99139
 A:Accession: C90506
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-543 <KUR>
 A:Cross-references: UNIPROT:Q97U16; GB:AE006641; NID:G13816647; PDB:AAK43306.1; GSPDB:G
 C:Genetics:
 A:Gene: ilvB-6

Query Match 88.2%; Score 30; DB 2; Length 543;
 Best Local Similarity 85.7%; Pred. No. 90;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELVR 7
 |:|||||
 Db 129 DSGELVR 135

RESULT 7
 S74372
 dnaK-type molecular chaperone dnaK - Synechocystis sp. (strain PCC 6803)
 N:Alternate names: protein sl10058
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 C:Accession: S74372
 R:Kanehisa, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;
 O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 s.
 A:Reference number: S74372; MUID:97061201; PMID:8905231
 A:Accession: S74372
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-692 <KAN>
 A:Cross-references: UNIPROT:Q55154; EMBL:D64001; GB:AB001339; NID:G1001102; PDB:BAA1029
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:32:16 ; Search time 14.5 Seconds
(without alignments)
46.449 Million cell updates/sec

Title: US-10-646-919-36

Perfect score: 34
Sequence: 1 DSGELVR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	91.2	123	2 F83507	hypothetical prote
2	31	91.2	299	2 B82164	hypothetical prote
3	31	91.2	300	2 T52379	zinc finger protei
4	31	91.2	1290	2 UC5473	dextranucrase (EC
5	30	88.2	235	2 T38909	hypothetical prote
6	30	88.2	543	2 C90506	hypothetical prote
7	30	88.2	692	2 S74372	dnak-type molecula
8	30	88.2	985	2 B86084	hypothetical prote
9	30	88.2	1377	2 B86034	rhba protein in rh
10	30	88.2	1377	2 C65159	rhba protein in rh
11	30	88.2	1394	2 H91236	rhsh core protein
12	30	88.2	1397	2 A85570	rhsc core protein
13	30	88.2	1397	2 C64805	rhsc core protein
14	30	88.2	1399	2 A99720	rhsc core protein
15	30	88.2	1409	2 F91187	rhba core protein
16	30	88.2	1411	2 B65145	rhba core protein
17	29	85.3	144	2 G70023	rhsh protein precu
18	29	85.3	157	2 F95313	hypothetical prote
19	29	85.3	173	2 A95864	hypothetical prote
20	29	85.3	178	2 T09584	high mobility grou
21	29	85.3	184	2 AE2330	adenylate kinase
22	29	85.3	304	2 C89908	hypothetical prote
23	29	85.3	312	2 B96512	hypothetical prote
24	29	85.3	420	2 S53324	glycogen synthase
25	29	85.3	420	1 TVRTKB	tau-protein kinase
26	29	85.3	449	2 S41950	utp-glucose glucos
27	29	85.3	453	2 C86176	hypothetical prote
28	29	85.3	467	2 D88710	protein C43G2.1 [l
29	29	85.3	645	2 F86627	lysr family transc

30	29	85.3	654	2 T32623	hypothetical prote
31	29	85.3	746	2 AF0845	hydrogenase matu
32	29	85.3	777	1 RPBP4	DNA primase - sate
33	29	85.3	777	2 C41830	DNA primase - phag
34	29	85.3	806	2 F69899	phage-related pre-
35	29	85.3	874	2 B70945	hypothetical prote
36	29	85.3	906	2 S35342	Goa1-associated p
37	29	85.3	906	2 S35312	coatomer complex b
38	29	85.3	954	2 T23369	hypothetical prote
39	29	85.3	1179	2 H82706	hypothetical prote
40	29	85.3	2541	2 S11661	talain - mouse
41	28	82.4	53	1 F25G1	photosystem II pro
42	28	82.4	101	2 B38189	protein-tyrosine-p
43	28	82.4	112	2 G86756	prophage p12 prote
44	28	82.4	173	2 B84348	hypothetical prote
45	28	82.4	255	2 G87677	conserved hypothet

ALIGNMENTS

RESULT 1
F83507
hypothetical protein PA1093 [imported] - Pseudomonas aeruginosa (strain PA01)
C/Spectes: Pseudomonas aeruginosa
C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C/Accession: F83507
R/Stover, C.K.; Pham, X.O.; Ewlin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.D.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Linn,
., Loty, S.; Olson, M.V.
Nature 406, 959-964, 2000
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A/Reference number: A82950; MUID:20437337; PMID:10984043
A/Accession: F83507
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-123 <STO>
A/Cross-references: UNIPROT:Q914N7; GB:AE004540; GB:AE004091; NID:G9947008; PIDN:AA6044
A/Experimental source: strain PA01
C/Genetic:
A/Genes: PA1093

Query Match 91.2%; Score 31; DB 2; Length 123;
Best Local Similarity 85.7%; Pred. No. 10;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
Db 91 DSGELVR 97

RESULT 2
B82164
hypothetical protein VC1738 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C/Spectes: Vibrio cholerae
C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C/Accession: B82164
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardon, D.; Ermolaeva, M.D.; Vamathavan, J.; Baer, S.; Qin, H.; Dragoi, I.; Sellers, F
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A/Reference number: A82035; MUID:20406833; PMID:10952301
A/Accession: B82164
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-259 <HEI>
A/Cross-references: UNIPROT:Q9KRA4; GB:AE004251; GB:AE003852; NID:G9656248; PIDN:AAF9488
A/Experimental source: serogroup O1, strain N16961, biotype El Tor
C/Genetic:
A/Genes: VC1738
A/Map position: 1
C/Superfamily: Xylella faecitiosa hypothetical protein XF1835

THIS PAGE IS BLANK

```
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangau
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITPA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43061
; LENGTH: 985
; TYPE: PRT
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (278)..(278)
; OTHER INFORMATION: X=any amino acid
US-10-282-122A-43061

Query Match      88.2%; Score 30; DB 15; Length 985;
Best Local Similarity 71.4%; Pred. No. 1.3e+03;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSGELVR 7
Db      452 DNGELVR 458

RESULT 15
US-10-425-115-312358
; Sequence 312358, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
```

```
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 312358
; LENGTH: 1017
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1017)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_47927C.1.pep
US-10-425-115-312358

Query Match      88.2%; Score 30; DB 16; Length 1017;
Best Local Similarity 85.7%; Pred. No. 1.4e+03;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 DSGELVR 7
Db      312 DSGELVR 318

Search completed: July 20, 2005, 14:32:01
Job time : 72.3333 secs
```

ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_40278C.1.pep
US-10-437-963-138936

Query Match 91.2%; Score 31; DB 16; Length 341;
Best Local Similarity 85.7%; Pred. No. 2.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 320 DSGELVR 326

RESULT 11
US-10-282-122A-47843
Sequence 47843, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Mang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreych, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
PRIOR FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 47843
LENGTH: 624
TYPE: PRT
ORGANISM: Burkholderia cepacia
US-10-282-122A-47843

Query Match 91.2%; Score 31; DB 15; Length 624;
Best Local Similarity 85.7%; Pred. No. 5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 317 DSGELVR 323

RESULT 12
US-10-369-493-2805
Sequence 2805, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 2805
LENGTH: 692
TYPE: PRT
ORGANISM: Synechocystis sp.
US-10-369-493-2805

Query Match 88.2%; Score 30; DB 15; Length 692;
Best Local Similarity 85.7%; Pred. No. 9.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 547 ESGELVR 553

RESULT 13
US-10-425-115-312359
Sequence 312359, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 312359
LENGTH: 976
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(976)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_47928C.1.pep
US-10-425-115-312359

Query Match 88.2%; Score 30; DB 16; Length 976;
Best Local Similarity 85.7%; Pred. No. 1.3e+03;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 312 DSGELVR 318

RESULT 14
US-10-282-122A-43061
Sequence 43061, Application US/10282122A
Publication No. US20040029129A1

```
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 140877
; LENGTH: 79
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_42033C.1.pep
US-10-437-963-140877
```

```
Query Match          91.2%; Score 31; DB 16; Length 79;
Best Local Similarity 85.7%; Pred. No. 56;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
        |||||
Db      35 DSGEVR 41
```

RESULT 7

```
US-09-864-408A-4672
; Sequence 4672, Application US/09864408A
; Publication No. US20040009474A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shinkets, Richard A.
; TITLE OF INVENTION: No. US20040009474A1 Human Polynucleotides and Polypeptides Encc
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4672
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-4672
```

```
Query Match          91.2%; Score 31; DB 11; Length 122;
Best Local Similarity 85.7%; Pred. No. 89;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
        |||||
Db      92 DNGELVR 98
```

RESULT 8

```
US-10-425-114-48472
; Sequence 48472, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
```

```
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 48472
; LENGTH: 323
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3150-008-C5_F11.pep
US-10-425-114-48472
```

```
Query Match          91.2%; Score 31; DB 15; Length 323;
Best Local Similarity 85.7%; Pred. No. 2,5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
        |||||
Db      300 DAGELVR 306
```

RESULT 9

```
US-10-425-115-319010
; Sequence 319010, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 319010
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_54009C.1.pep
US-10-425-115-319010
```

```
Query Match          91.2%; Score 31; DB 16; Length 329;
Best Local Similarity 85.7%; Pred. No. 2,5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
        |||||
Db      307 DAGELVR 313
```

RESULT 10

```
US-10-437-963-138936
; Sequence 138936, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 138936
; LENGTH: 341
; TYPE: PRT
```



```
RESULT 2
US-10-425-115-351857
; Sequence 351857, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; LENGTH: 113
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_84058C.1.pep
US-10-425-115-351857

Query Match
Best Local Similarity 100.0%; Score 34; DB 16; Length 113;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 96 DSGELVR 102

RESULT 3
US-10-425-114-46116
; Sequence 46116, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46116
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 701179796_FLI.pep
US-10-425-114-46116

Query Match
Best Local Similarity 100.0%; Score 34; DB 15; Length 125;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 108 DSGELVR 114

RESULT 4
US-10-425-115-351854
; Sequence 351854, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1)..(214)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_84055C.1.pep
US-10-425-115-351854

Query Match
Best Local Similarity 100.0%; Score 34; DB 16; Length 214;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 97 DSGELVR 103

RESULT 5
US-10-646-919-35
; Sequence 35, Application US/10646919
; Publication No. US20050148075A1
; GENERAL INFORMATION:
; APPLICANT: Barbas, Carlos F.
; TITLE OF INVENTION: Zinc-finger binding domains for GNN
; FILE REFERENCE: TSRI 645.2c1
; CURRENT APPLICATION NUMBER: US/10/646,919
; CURRENT FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: US 09/494,190
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/EP99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthesized
US-10-646-919-35

Query Match
Best Local Similarity 91.2%; Score 31; DB 18; Length 7;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 1 DSGELVR 7

RESULT 6
US-10-437-963-140877
; Sequence 140877, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1)..(214)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_84055C.1.pep
US-10-425-115-351854

Query Match
Best Local Similarity 100.0%; Score 34; DB 16; Length 113;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 96 DSGELVR 102

RESULT 3
US-10-425-114-46116
; Sequence 46116, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46116
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 701179796_FLI.pep
US-10-425-114-46116

Query Match
Best Local Similarity 100.0%; Score 34; DB 15; Length 125;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 108 DSGELVR 114

RESULT 4
US-10-425-115-351854
; Sequence 351854, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1)..(214)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_84055C.1.pep
US-10-425-115-351854

Query Match
Best Local Similarity 100.0%; Score 34; DB 16; Length 214;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 97 DSGELVR 103

RESULT 5
US-10-646-919-35
; Sequence 35, Application US/10646919
; Publication No. US20050148075A1
; GENERAL INFORMATION:
; APPLICANT: Barbas, Carlos F.
; TITLE OF INVENTION: Zinc-finger binding domains for GNN
; FILE REFERENCE: TSRI 645.2c1
; CURRENT APPLICATION NUMBER: US/10/646,919
; CURRENT FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: US 09/494,190
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/EP99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthesized
US-10-646-919-35

Query Match
Best Local Similarity 91.2%; Score 31; DB 18; Length 7;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 1 DSGELVR 7

RESULT 6
US-10-437-963-140877
; Sequence 140877, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; LENGTH: 214
; TYPE: PRT
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1)..(214)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_84055C.1.pep
US-10-425-115-351854

Query Match
Best Local Similarity 100.0%; Score 34; DB 16; Length 214;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 97 DSGELVR 103
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:52:31 ; Search time 71.3333 Seconds
(without alignments)
38.093 Million cell updates/sec

Title: US-10-646-919-36
Perfect score: 34
Sequence: 1 DSGELVR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1736639 seqs, 388188149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	100.0	7	18	US-10-646-919-36
2	34	100.0	113	16	US-10-425-115-351857
3	34	100.0	125	15	US-10-425-114-46116
4	34	100.0	214	16	US-10-425-115-351854
5	31	91.2	7	18	US-10-646-919-35
6	31	91.2	79	16	US-10-437-963-140877
7	31	91.2	122	11	US-09-864-408A-4672
8	31	91.2	323	15	US-10-425-114-48472
9	31	91.2	329	16	US-10-425-115-319010
10	31	91.2	341	16	US-10-437-963-138936
11	31	91.2	624	15	US-10-282-122A-47843

	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45																																																																																																																																																																
	30	88.2	692	15	US-10-369-493-2805	30	88.2	976	16	US-10-425-115-312359	30	88.2	985	15	US-10-282-122A-43061	30	88.2	1017	16	US-10-425-115-312358	30	88.2	1377	9	US-09-815-242-10384	30	88.2	1377	14	US-10-287-274-467	30	88.2	1377	15	US-10-282-122A-42731	30	88.2	1397	15	US-10-282-122A-43059	30	88.2	1411	15	US-10-282-122A-43060	30	88.2	5183	14	US-10-107-521-1	22	29	85.3	44	9	US-09-095-881-15	22	29	85.3	60	15	US-10-424-559-315210	24	29	85.3	88	16	US-10-425-115-330300	25	29	85.3	163	16	US-10-437-963-114359	26	29	85.3	167	16	US-10-437-963-105651	27	29	85.3	178	17	US-10-733-923-12994	28	29	85.3	190	16	US-10-437-963-140502	29	29	85.3	253	14	US-10-238-075-1268	30	29	85.3	259	16	US-10-437-963-178622	31	29	85.3	304	15	US-10-282-122A-44419	32	29	85.3	308	9	US-09-815-242-5258	33	29	85.3	319	9	US-09-815-242-12312	34	29	85.3	320	16	US-10-425-115-256929	35	29	85.3	345	16	US-10-437-963-114562	36	29	85.3	350	16	US-10-746-545-27	37	29	85.3	352	16	US-10-664-421-31	38	29	85.3	352	16	US-10-746-545-16	39	29	85.3	352	16	US-10-746-545-17	40	29	85.3	361	9	US-09-916-109-3	41	29	85.3	361	14	US-10-211-412A-3	42	29	85.3	361	16	US-10-450-422-1	43	29	85.3	361	17	US-10-689-461-3	44	29	85.3	367	16	US-10-746-545-32	45	29	85.3	385	16	US-10-664-421-53

ALIGNMENTS

RESULT 1
US-10-646-919-36
; Sequence 36, Application US/10646919
; Publication No. US20050148075A1
; GENERAL INFORMATION:
; APPLICANT: Barbae, Carlos F.
; TITLE OF INVENTION: Zinc-finger binding domains for GNN
; FILE REFERENCE: TSRI 645.2C1
; CURRENT APPLICATION NUMBER: US/10/646,919
; PRIOR FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: US 09/494,190
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/EP99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthesized
US-10-646-919-36

Query Match 100.0%; Score 34; DB 18; Length 7;
Best local Similarity 100.0%; Pred. No. 1.6e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
DB 1 DSGELVR 7

```
FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:phage display
; OTHER INFORMATION: selected and mutagenized
US-09-494-190-36
```

```
Query Match      100.0%; Score 34; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
        |||||
Db       1 DSGDLVR 7
```

```
RESULT 3
US-09-173-941-35
; Sequence 35, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV00815
; CURRENT APPLICATION NUMBER: US/09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
US-09-173-941-35
```

```
Query Match      91.2%; Score 31; DB 3; Length 7;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
        |||||
Db       1 DSGDLVR 7
```

```
RESULT 4
US-09-494-190-35
; Sequence 35, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 35
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:phage display
US-09-494-190-35
```

```
Query Match      91.2%; Score 31; DB 4; Length 7;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
        |||||
Db       1 DSGDLVR 7
```

```
RESULT 5
US-09-711-164-467
; Sequence 467, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Foreyth, R. Allyn
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERET
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 467
; LENGTH: 1377
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-711-164-467
```

```
Query Match      88.2%; Score 30; DB 4; Length 1377;
Best Local Similarity 71.4%; Pred. No. 4.7e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELVR 7
        |||||
Db       861 DNGELIR 867
```

```
RESULT 6
US-09-095-881-15
; Sequence 15, Application US/09095881
; Patent No. 6469137
; GENERAL INFORMATION:
; APPLICANT: Seeley, Todd
; TITLE OF INVENTION: DETECTION OF LOSS OF THE WILD-TYPE
; FILE REFERENCE: 1405.003 / 200130.437
; CURRENT APPLICATION NUMBER: US/09/095,881
; PRIOR FILING DATE: 1998-06-11
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-095-881-15
```

```
Query Match      85.3%; Score 29; DB 4; Length 44;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 DSGELV 6
        |||||
Db       24 DSGELV 29
```

```
RESULT 7
US-09-489-039A-8264
; Sequence 8264, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Bretton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:17:05 ; Search time 21.6667 Seconds
(without alignments)
24.117 Million cell updates/sec

Title: US-10-646-919-36

Perfect score: 34

Sequence: 1 DSGELVR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/CCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/backfillseq.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	34	100.0	7	3	US-09-173-941-36	Sequence 36, Appl
2	34	100.0	7	4	US-09-494-190-36	Sequence 36, Appl
3	31	91.2	7	3	US-09-173-941-35	Sequence 35, Appl
4	31	91.2	7	4	US-09-494-190-35	Sequence 35, Appl
5	30	88.2	1377	4	US-09-711-164-467	Sequence 467, Appl
6	29	85.3	44	4	US-09-095-881-15	Sequence 15, Appl
7	29	85.3	304	4	US-09-489-039A-8264	Sequence 8264, Ap
8	29	85.3	354	4	US-09-540-236-2689	Sequence 2689, Ap
9	29	85.3	361	4	US-09-916-109-3	Sequence 3, Appl
10	29	85.3	361	4	US-10-211-412B-3	Sequence 3, Appl
11	29	85.3	394	4	US-09-916-109-2	Sequence 2, Appl
12	29	85.3	394	4	US-10-211-412B-2	Sequence 2, Appl
13	29	85.3	420	2	US-08-602-264A-14	Sequence 14, Appl
14	29	85.3	420	4	US-09-316-038-1	Sequence 1, Appl
15	29	85.3	420	4	US-09-916-109-1	Sequence 1, Appl
16	29	85.3	420	4	US-10-211-412B-1	Sequence 1, Appl
17	29	85.3	420	4	US-09-538-092-1163	Sequence 1163, Ap
18	29	85.3	425	4	US-09-270-767-45473	Sequence 45473, A
19	29	85.3	897	4	US-09-802-540-11551	Sequence 11551, A
20	29	85.3	905	4	US-09-538-092-1079	Sequence 1079, Ap
21	29	85.3	906	1	US-08-190-802A-31	Sequence 31, Appl
22	29	85.3	906	1	US-08-477-346-31	Sequence 31, Appl
23	29	85.3	906	3	US-08-473-089-31	Sequence 31, Appl
24	29	85.3	906	4	US-08-487-072A-31	Sequence 31, Appl
25	28	82.4	7	3	US-09-173-941-8	Sequence 8, Appl
26	28	82.4	7	3	US-09-173-941-37	Sequence 37, Appl
27	28	82.4	7	3	US-09-173-941-77	Sequence 77, Appl

28	28	82.4	7	3	US-09-173-941-107	Sequence 107, App
29	28	82.4	7	4	US-09-494-190-8	Sequence 8, Appl
30	28	82.4	7	4	US-09-494-190-37	Sequence 37, Appl
31	28	82.4	7	4	US-09-494-190-77	Sequence 77, Appl
32	28	82.4	7	4	US-09-494-190-107	Sequence 107, App
33	28	82.4	101	4	US-09-134-000C-6643	Sequence 6643, Ap
34	28	82.4	115	3	US-09-303-120B-4	Sequence 4, Appl
35	28	82.4	115	3	US-09-820-576-4	Sequence 4, Appl
36	28	82.4	115	4	US-09-966-608-4	Sequence 4, Appl
37	28	82.4	171	4	US-09-270-767-34902	Sequence 34902, A
38	28	82.4	171	4	US-09-270-767-50119	Sequence 50119, A
39	28	82.4	197	4	US-09-107-532A-5430	Sequence 5430, Ap
40	28	82.4	202	3	US-09-171-461-46	Sequence 46, Appl
41	28	82.4	202	4	US-09-970-711-46	Sequence 46, Appl
42	28	82.4	256	4	US-09-252-991A-21603	Sequence 21603, A
43	28	82.4	281	1	US-08-487-748A-9	Sequence 9, Appl
44	28	82.4	281	1	US-08-487-748A-10	Sequence 10, Appl
45	28	82.4	281	3	US-08-480-070C-10	Sequence 10, Appl

ALIGNMENTS

```
RESULT 1
US-09-173-941-36
; Sequence 36, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV00815
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
; OTHER INFORMATION: codon binding sequence
US-09-173-941-36

Query Match      100.0%; Score 34; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 4; le+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 DSGELVR 7
Db      1 DSGELVR 7

RESULT 2
US-09-494-190-36
; Sequence 36, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

ID ADJ98426 standard; peptide; 7 AA.
 XX ADJ98426;
 AC
 XX
 DT 06-MAY-2004 (first entry)
 XX
 DE Zinc finger DNA binding peptide #97.
 XX
 KM library; multimeric DNA binding polypeptide;
 KM zinc finger DNA binding peptide; gene expression silencing;
 KM gene expression enhancement.
 XX
 OS Unidentified.
 XX
 PN WO2003066828-A2.
 XX
 PD 14-AUG-2003.
 XX
 PF 07-FEB-2003; 2003WO-US003705.
 XX
 PR 07-FEB-2002; 2002US-0354981P.
 XX
 PA (SCR1) SCRIPPS RES INST.
 XX
 PI Barbas CF, Blancafort P;
 XX
 DR WPI; 2003-731499/69.
 XX
 PT New zinc finger library of multimeric DNA binding polypeptides, useful
 PT for sterically occluding the binding site of a natural transcription
 PT factor, and enhancing or silencing target gene expression.
 XX
 PS Disclosure; Fig 20; 64pp; English.
 XX
 CC The invention comprises a library of multimeric DNA binding polypeptides
 CC (e.g. zinc finger DNA binding polypeptides). The zinc finger binding
 CC polypeptides of the invention are useful for sterically occluding the
 CC binding site of a natural transcription factor and enhancing or silencing
 CC target gene expression. The present amino acid sequence represents a zinc
 CC finger DNA binding peptide of the invention.
 CC
 SQ Sequence 7 AA;
 XX
 QY Query Match 91.2%; Score 31; DB 7; Length 7;
 Db Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 1 DSGELVR 7
 1 DSGDLYR 7
 RESULT 7
 ABP33363
 ID ABP33363 standard; protein; 122 AA.
 XX
 AC ABP33363;
 XX
 DT 08-JUL-2002 (first entry)
 XX
 DE Human ORF2336 protein, SEQ ID NO:4672.
 XX
 KM Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
 KM disease monitoring; cytokine; cell proliferation; cell differentiation;
 KM immune modulation; haematopoiesis regulation; tissue growth;
 KM angiogenesis; actinin; inhibitor; chemotactic; chemokinetic; haemostatic;
 KM chromolytic; tumour inhibition; bodily characteristics; fertility;
 KM behaviour; cancer; proliferative disorder; neurological disorder;
 KM cardiovascular disease; immune system disorder; organ transplantation;
 KM tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
 KM hypothyroidism; cholesterol ester storage disease; infection; vulnerability;
 KM vasoconstrictive; antiproliferative; antidiabetic; cytostatic; neurotropic;
 KM neuroprotective; antithrombotic; anticoagulant; thrombolytic;

KM cardiant; hypotensive; antihydroid; antiinflammatory; immunomodulator;
 KM dermatological; analgesic; virucide; antibacterial; fungicide.
 XX
 OS Homo sapiens.
 XX
 PN WO200190366-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 24-MAY-2001; 2001WO-US017076.
 XX
 PR 24-MAY-2000; 2000US-0206690P.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Leach MD, Shinkels RA;
 XX
 DR WPI; 2002-106200/14.
 XX
 DR N-PDSB; ABN77389.
 XX
 PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and disorders related to organ
 PT transplantation.
 XX
 PS Claim 10; Page 1421; 2508pp; English.
 XX
 CC Sequences ABP31028-ABP3561 represent 4534 novel human proteins
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
 CC polypeptides, methods of screening for modulators of ORFX expression or
 CC activity, and methods of screening individuals for a predisposition to an
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
 CC range of biological activities, such as cytokine, cell proliferation,
 CC cell differentiation, immune modulation, haematopoiesis regulation,
 CC tissue growth, angiogenesis, actinin or inhibin activity, chemotactic/
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
 CC and antiinfective activity, and may also be involved in the determination
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,
 CC nucleic acids and antibodies may be used in the treatment of cancers,
 CC other proliferative disorders such as psoriasis and benign tumours,
 CC neurological disorders such as epilepsy and Alzheimer's disease,
 CC cardiovascular diseases, immune system disorders, disorders related to
 CC organ transplantation, disorders of tissue growth and regeneration,
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
 CC storage disease, and infectious diseases caused by viral, bacterial,
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a
 CC source of primers and probes, in the detection of ORFX genomic sequences
 CC or transcripts, in the identification and cloning of homologous
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
 CC nucleic acids may additionally be used to produce transgenic animals
 CC which may be useful for studying the function and/or activity of ORFX
 CC protein, and in drug screening. The ORFX proteins may also be used as
 CC immunogens to generate specific antibodies, which are useful in the
 CC diagnosis, treatment and monitoring of ORFX-associated diseases
 CC
 SQ Sequence 122 AA;
 XX
 QY Query Match 91.2%; Score 31; DB 5; Length 122;
 Db Best Local Similarity 85.7%; Pred. No. 65;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 1 DSGELVR 7
 1 DSGDLYR 98

RESULT 4
AAB02894
ID AAB02894 standard; peptide; 7 AA.
XX
AC AAB02894;
XX
DT 18-SEP-2000 (first entry)
XX
DE Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:35.
XX
KM zinc finger domain; alpha helix; nucleotide binding; DNA binding;
KW polydactyl protein; asymmetric target recognition;
KW gene specific transcriptional regulator; gene activator; gene repressor;
KM transcriptional switch; oncogene; erbB-2; cancer; tumour; gene therapy;
KM transgenic animal; antiviral; anticancer; diagnosis.
XX
OS Synthetic.
XX
PN WO200023464-A2.
XX
PD 27-APR-2000.
XX
PF 14-OCT-1999; 99WO-EP007742.
XX
PR 16-OCT-1998; 98US-00173941.
XX
PA (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PA (SCRI) SCRIPPS RES INST.
XX
PI Barbas CF;
XX
DR MPI; 2000-339648/29.
XX
PT Novel isolated and purified zinc finger nucleotide-binding proteins with
PT specificity for GNN triplet sequences, useful in gene therapy and for
PT regulating gene function.
XX
PS Disclosure; Fig 1; 48pp; English.
XX
CC The invention relates to zinc finger nucleotide-binding proteins which
CC comprise 2-12, preferably 2-6, operatively linked motifs selected from
CC sequences AAB02860-B02875. Sequences AAB02860-B02875 represent the alpha
CC helical regions of zinc finger domains which specifically bind to target
CC nucleotide triplets of the sequence 5'-GNN-3'. Such regions may be linked
CC by the peptide linker TGRKP (AAB02970). The Cys2-His2 zinc finger motif
CC is the most frequently utilised nucleic acid binding motif in eukaryotes,
CC and constitutes a beta-beta-alpha fold. Nucleic acid recognition is
CC achieved through specific contacts from side chains of amino acid
CC residues in the alpha helix. Each zinc finger can recognise a sub-site of
CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains
CC allows the recognition of extended contiguous asymmetric DNA sequences.
CC For example, a synthetic polydactyl protein containing six zinc finger
CC domains can recognise an 18 bp sequence, and such proteins are
CC potentially highly gene-specific. The novel nucleotide-binding zinc
CC finger proteins may therefore be used in the development of artificial
CC gene-specific transcriptional regulators. Such transcriptional switches
CC may be used to regulate the expression of oncogenes such as erbB-2,
CC overexpression of which is involved in malignant transformation. The
CC proteins are therefore useful in the treatment of cancers, and may also
CC be used to activate genes involved in fighting diseases, and to treat
CC viral infections by inhibiting the synthesis of viral gene products. They
CC may be used in DNA-based diagnostic applications. The proteins may also
CC be used in producing functional gene knockout or activation in
CC heterozygous transgenic animals. Proteins of the invention can
CC discriminate between sequences which have a single base difference.
CC This is manifested in a >100-fold decrease in affinity for the variant
CC sequence. Gene activation and repression can be achieved by targeting
CC within the gene transcript, suggesting that information obtained from
CC expressed sequence tags may be sufficient for the construction of gene
CC switches. Sequences AAB02876-B02869 represent zinc finger alpha helix

CC phase library peptides disclosed in the invention
XX
SQ Sequence 7 AA;
XX
Query Match 91.2%; Score 31; DB 3; Length 7;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 DSGELVR 7
|||:||||
Db 1 DSGDLVR 7
XX
RESULT 5
ADFe1984
ID ADFe1984 standard; peptide; 7 AA.
XX
AC ADFe1984;
XX
DT 12-FEB-2004 (first entry)
XX
DE Zinc finger binding region #35.
XX
KM zinc finger-nucleotide binding polypeptide; expression regulation;
KW zinc finger binding region.
XX
OS Synthetic.
XX
PN US6610512-B1.
XX
PD 26-AUG-2003.
XX
PF 28-JAN-2000; 2000US-00494190.
XX
PR 16-OCT-1998; 98US-00173941.
PR 14-OCT-1999; 99WO-EP007742.
XX
PA (SCRI) SCRIPPS RES INST.
XX
PI Barbas CF;
XX
DR MPI; 2003-800134/75.
XX
PT Regulating expression of nucleotide sequence that contains sequence 5'-
PT (GNN)n-3', comprises exposing nucleotide sequence to composition
PT comprising isolated and purified zinc finger-nucleotide binding
PT polypeptide.
XX
PS Disclosure; SEQ ID NO 35; 46pp; English.
XX
CC The invention relates to an expression of nucleotide sequence that
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by
CC exposing nucleotide sequence to composition comprising isolated and
CC purified zinc finger-nucleotide binding polypeptide that contains a
CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Leu-Val-
CC Arg. The invention is used in the regulation of the expression of the
CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present
CC sequence represents the amino acid sequence of a zinc finger binding
CC region.
XX
SQ Sequence 7 AA;
XX
Query Match 91.2%; Score 31; DB 7; Length 7;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 DSGELVR 7
|||:||||
Db 1 DSGDLVR 7
XX
RESULT 6
ADJ98426

CC achieved through specific contacts from side chains of amino acid
CC residues in the alpha helix. Each zinc finger can recognise a subside of
CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains
CC allows the recognition of extended contiguous asymmetric DNA sequences.
CC For example, a synthetic polypeptide protein containing six zinc finger
CC domains can recognise an 18 bp sequence, and such proteins are
CC potentially highly gene-specific. The novel nucleotide-binding zinc
CC finger proteins may therefore be used in the development of artificial
CC gene-specific transcriptional regulators. Such transcriptional switches
CC may be used to regulate the expression of oncogenes such as erbB-2,
CC overexpression of which is involved in malignant transformation. The
CC proteins are therefore useful in the treatment of cancers, and may also
CC be used to activate genes involved in fighting diseases, and to treat
CC viral infections by inhibiting the synthesis of viral gene products. They
CC may be used in DNA-based diagnostic applications. The proteins may also
CC be used in producing functional gene knockout or activation in
CC heterozygous transgenic animals. Proteins of the invention can
CC discriminate between sequences which have a single base difference. This
CC is manifested in a >100-fold decrease in affinity for the variant
CC sequence. Gene activation and repression can be achieved by targeting
CC within the gene transcript, suggesting that information obtained from
CC expressed sequence tags may be sufficient for the construction of gene
CC switches. Sequences AAB02876-B02869 represent zinc finger alpha helix
CC phage library peptides disclosed in the invention
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 34; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
| | | | |
Db 1 DSGELVR 7

RESULT 2

ADF61985 ADF61985 standard; peptide; 7 AA.

XX ADF61985;

DT 12-FEB-2004 (first entry)

XX Zinc finger binding region #36.

XX zinc finger-nucleotide binding polypeptide; expression regulation;

XX zinc finger binding region.

OS Synthetic.

PN US6610512-B1.

PD 26-AUG-2003.

PF 28-JUN-2000; 2000US-00494190.

PR 16-OCT-1998; 98US-00173941.

PR 14-OCT-1999; 99WO-EP007742.

PA (SCRI) SCRIPPS RES INST.

PI Barbas CF;

XX WPI; 2003-800134/75.

PT Regulating expression of nucleotide sequence that contains sequence 5'-
PT (GNN)n-3', comprises exposing nucleotide sequence to composition
PT comprising isolated and purified zinc finger-nucleotide binding
PT polypeptide.

PS Disclosure; SEQ ID NO.36; 46pp; English.
XX

CC The invention relates to an expression of nucleotide sequence that
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by
CC exposing nucleotide sequence to composition comprising isolated and
CC purified zinc finger-nucleotide binding polypeptide that contains a
CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Ieu-Val-
CC Arg. The invention is used in the regulation of the expression of the
CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present
CC sequence represents the amino acid sequence of a zinc finger binding
CC region.
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 34; DB 7; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
| | | | |
Db 1 DSGELVR 7

RESULT 3

ADJ98427 ADJ98427 standard; peptide; 7 AA.

XX ADJ98427;

DT 06-MAY-2004 (first entry)

DE Zinc finger DNA binding peptide #98.

XX library; multimeric DNA binding polypeptide;

XX zinc finger DNA binding peptide; gene expression silencing;

XX gene expression enhancement.

XX Unidentified.

PN WO2003066828-A2.

PD 14-AUG-2003.

PF 07-FEB-2003; 2003WO-US003705.

PR 07-FEB-2002; 2002US-0354981P.

PA (SCRI) SCRIPPS RES INST.

PI Barbas CF, Blancafort P;

XX WPI; 2003-731499/69.

PT New zinc finger library of multimeric DNA binding polypeptides, useful
PT for sterically occluding the binding site of a natural transcription
PT factor, and enhancing or silencing target gene expression.

PS Disclosure; Fig 20; 64pp; English.
XX

CC The invention comprises a library of multimeric DNA binding polypeptides
CC (e.g. zinc finger DNA binding polypeptides). The zinc finger binding
CC polypeptides of the invention are useful for sterically occluding the
CC binding site of a natural transcription factor and enhancing or silencing
CC target gene expression. The present amino acid sequence represents a zinc
CC finger DNA binding peptide of the invention.
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 34; DB 7; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGELVR 7
| | | | |
Db 1 DSGELVR 7

Query Match 91.4%; Score 32; DB 2; Length 160;
 Best Local Similarity 85.7%; Pred. No. 88;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
 DB 113 DSGDLVR 119

RESULT 8
 ID 06LBC7 PRELIMINARY; PRT; 167 AA.
 AC 06LBC7;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Hypothetical protein.
 OS Oligotropha carboxidovorans (Pseudomonas carboxydovorans).
 OC Plasmid pHC3.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Oligotropha.
 OX NCBI_TaxID=40137;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OMS;
 RX MEDLINE=95238294; PubMed=7721710;
 RA Schubel U., Kraut M., Moersdorf G., Meyer O.;
 RT "Molecular characterization of the gene cluster coxMSL encoding the
 RT molycopium-containing carbon monoxide dehydrogenase of Oligotropha
 RT carboxidovorans.";
 RL J. Bacteriol. 177:2197-2197(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OMS;
 RX MEDLINE=97464431; PubMed=9324252;
 RA Santiago B., Meyer O.;
 RT "Purification and molecular characterization of the H2 uptake
 RT membrane-bound Nite-hydrogenase from the carboxidotrophic bacterium
 RT Oligotropha carboxidovorans.";
 RL J. Bacteriol. 179:6053-6060(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OMS;
 RA Santiago B., Schubel U., Egeleser C., Meyer O.;
 RT "Sequence analysis, characterization and CO-specific transcription of
 RT the cox gene cluster on the megaplasmid pHC3 of Oligotropha
 RT carboxidovorans.";
 RL Gene 236:1157-1247(1999).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OMS;
 RX PubMed=14644498; DOI=10.1016/j.gene.2003.08.027;
 RA Fuhrmann S., Farmer M., Jeffke T., Henne A., Gottschalk G., Meyer O.;
 RT "Complete nucleotide sequence of the self-transmissible circular
 RT megaplasmid pHC3 of Oligotropha carboxidovorans. Function in the
 RT chemolithoautotrophic utilization of CO, H2 and CO2.";
 RL Gene 322:67-75(2003).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OMS;
 RA Schubel U.;
 RT Submitted (JUL-1993) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OMS;
 RX PubMed=14644498; DOI=10.1016/j.gene.2003.08.027;
 RA Fuhrmann S.;
 RT Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X62447; CAG28432.1; -;
 KW Hypothetical protein; Plasmid.
 RT SEQUENCE 167 AA; 18757 MW; F6702FF2A6A55553 CRC64;
 Query Match 91.4%; Score 32; DB 2; Length 167;

Best Local Similarity 85.7%; Pred. No. 93;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
 DB 10 DSGDLVR 16

RESULT 9
 ID 08R5HO PRELIMINARY; PRT; 313 AA.
 AC 08R5HO;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE SH2 phosphatase 1 (Fragment).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mistar;
 RA Heneberg P., Draber P.;
 RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF468653; AAL77056.1; -;
 DR HSSP; P29350; 1GW2.
 DR GO; GO:0016787; F:hydrolase activity; IEA.
 DR GO; GO:0004725; P:protein tyrosine phosphatase activity; IEA.
 DR GO; GO:0004770; P:protein amino acid dephosphorylation; IEA.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR InterPro; IPR000242; TYR_PP.
 DR Pfam; PF00102; Y_phosphatase; 1.
 DR PRINTS; PR00700; PRTPHPTASE.
 DR SMART; SM00194; PTPc; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
 KW Hydrolase.
 FT NON_TER
 QY SEQUENCE 313 AA; 35605 MW; B8A1085FE7737C8B CRC64;
 Query Match 91.4%; Score 32; DB 2; Length 313;
 Best Local Similarity 85.7%; Pred. No. 19e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
 DB 119 DSGDLVR 125

RESULT 10
 ID 07UM19 PRELIMINARY; PRT; 452 AA.
 AC 07UM19;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Signal-transducing histidine kinase (EC 2.7.3.11).
 GN OrderedCusNames=RB2008;
 OS Rhodospirillum rubrum.
 OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
 OC Planctomycetaceae; Pirellula.
 OX NCBI_TaxID=117;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1;
 RX MEDLINE=22735913; PubMed=12835416; DOI=10.1073/pnas.1431443100;
 RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
 RA Ludwig W., Gade D., Beck A., Borzym K., Heilmann K., Rabus R.,
 RA Schlesner H., Amann R., Reinhardt R.;
 RT "Complete genome sequence of the marine planctomycete Pirellula sp.
 RT strain 1.";

OC Phyllobacteriaceae; Mesorhizobium.
 OK NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RT DNA Res. 7:331-338(2000).
 RL EMBL, AP003013; BAB53758.1; -.
 DR InterPro; IPR008940; Preamyl_trans.
 DR InterPro; IPR001440; TPR.
 DR SMART; SM00026; TPR; 4.
 DR PROSITE; PSS0005; TPR; 2.
 DR PROSITE; PSS0293; TPR_REGION; 2.
 KW Complete proteome.
 SQ SEQUENCE 289 AA; 3086 MW; D8F9912BD4E87111 CRC64;

Query Match 97.1%; Score 34; DB 2; Length 289;
 Best Local Similarity 85.7%; Pred. No. 63;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DSGDLVR 7
 Db 8 DSGDLIR 14

RESULT 5

07RGE6 PRELIMINARY; PRT; 400 AA.
 AC 07RGE6;
 DT 01-MAR-2004 (TRMBLrel. 26, Created)
 DT 01-MAR-2004 (TRMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TRMBLrel. 26, Last annotation update)
 DE Hypothetical protein (Fragment).
 GN Name=PY04401;
 OS Plasmodium yoelii yoelii;
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OK NCBI_TaxID=73239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=17XNL;
 RX PubMed=12368865; DOI=10.1038/nature01099;
 RA Carlton J.M., Anguilo S.V., Suh B.B., Kooij T.W., Pertea M.,
 Silva J.C., Ermolaeva M.D., Allen J.E., Sengen J.D., Koo H.L.,
 Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,
 RA Shallom S.J., van Aken S.E., Riedmiller S.B., Feldlynn T.V.,
 Cho J.K., Queckenbush J., Sedegah M., Shoat A., Cummings L.M.,
 Florens L., Yates F.R., III, Raine J.D., Sinden R.E., Harris M.A.,
 Cunningham D.A., Preiser P.R., Bergman L.W., Vaideya A.B.,
 van Lin L.H., Jance C.J., Waters A.P., Smith H.O., White O.R.,
 RA Salberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
 RA Carucci D.J.;
 RT "Genome sequence and comparative analysis of the model rodent malaria
 parasite Plasmodium yoelii yoelii.";
 RT Nature 419:512-519(2002).
 CC -1- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AABL0100133; EAA16258.1; -.
 KW Hypothetical protein.
 RT NON TER 400
 FT 400
 SQ SEQUENCE 400 AA; 47117 MW; DD2FB28296B0303A CRC64;

Query Match 94.3%; Score 33; DB 2; Length 400;
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DSGDLVR 7
 Db 102 DSGDLVR 108

RESULT 6

090U6 PRELIMINARY; PRT; 101 AA.
 ID 090U6;
 AC 090U6;
 DT 01-MAY-2000 (TRMBLrel. 13, Created)
 DT 01-MAY-2000 (TRMBLrel. 13, Last sequence update)
 DT 01-JUN-2003 (TRMBLrel. 24, Last annotation update)
 DE Tyrosine phosphatase from CDNA clone SH-PTP1 (Fragment).
 OS Rattus sp.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OK NCBI_TaxID=10118;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92141214; PubMed=1736296;
 RA Plutsky J., Neel B.G., Rosenberg R.D.;
 RT "Isolation of a src homology 2-containing tyrosine phosphatase.";
 RT Proc. Natl. Acad. Sci. U.S.A. 89:1123-1127(1992).
 DR HSBP; P29350; IGWZ.
 DR GO; GO:0004725; F:protein tyrosine phosphatase activity; IEA.
 DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
 DR InterPro; IPR000387; P:protein amino acid dephosphorylation; IEA.
 DR InterPro; IPR00242; Tyr_PP.
 DR Pfam; PF00102; Y_phosphatase; 1.
 DR PRINTS; PR00700; PRTYPHRTASE.
 DR PROSITE; PSS0056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PSS0055; TYR_PHOSPHATASE_PTP; 1.
 FT NON TER 101
 FT 101
 SQ SEQUENCE 101 AA; 11404 MW; 04FAF62531D5A3 CRC64;

Query Match 91.4%; Score 32; DB 2; Length 101;
 Best Local Similarity 85.7%; Pred. No. 63;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DSGDLVR 7
 Db 42 DNGDLVR 48

RESULT 7

098304 PRELIMINARY; PRT; 160 AA.
 ID 098304;
 AC 098304;
 DT 01-OCT-2001 (TRMBLrel. 18, Created)
 DT 01-OCT-2001 (TRMBLrel. 18, Last sequence update)
 DT 01-MAR-2002 (TRMBLrel. 20, Last annotation update)
 DE M18225 protein.
 GN Ordered locus names=ml8225;
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Phyllobacteriaceae; Mesorhizobium.
 OK NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,
 Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsumoto A.,
 Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RT DNA Res. 7:331-338(2000).
 RL EMBL, AP003013; BAB53826.1; -.
 KW Complete proteome.
 SQ SEQUENCE 160 AA; 16473 MW; 2F352124BDEDA641 CRC64;

OK NCBI_TaxId=92829;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=306 / ATCC 13902 / XV 101;
 RX MEDLINE=2202245; PubMed=12024217; DOI=10.1038/417459a;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Clapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Fomighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Melandri J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.;
 RA "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities";
 RL Nature 417:459-463 (2002).
 CC -1- FUNCTION: Cell wall formation.
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmutamate + L-alanine = ADP +
 CC phosphate + UDP-N-acetylmutamoyl-L-alanine.
 CC -1- PATHWAY: Peptidoglycan biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: Belongs to the murCDP family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AE011708; AM35668.1; -;
 CC DR HSSP; P45066; IG00.
 CC DR HAMAP; MF_00046; -; 1.
 CC DR InterPro; IPR004101; Mur_1lgase_C.
 CC DR InterPro; IPR000713; Mur_1lgase_N.
 CC DR InterPro; IPR005758; MurC.
 CC DR Pfam; PF01225; Mur_1lgase; 1.
 CC DR Pfam; PF02875; Mur_1lgase; 1.
 CC DR TIGRFAMs; TIGR01082; murC; 1.
 CC DR ATP-binding; Cell division; Cell wall; Complete proteome; Ligase;
 CC KW NP_BIND 122 128 ATP (potential).
 CC FT SEQUENCE 477 AA; 50592 MW; 53F60B180C71A2F CRC64;
 CC SQ
 Query Match 100.0%; Score 35; DB 1; Length 477;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 DSGDLVR 7
 Db 7 DSGDLVR 13
 RESULT 3
 MURC_XANCP STANDARD; PRT; 477 AA.
 ID MURC_XANCP
 AC 08PC39;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE UDP-N-acetylmutamate--L-alanine ligase (EC 6.3.2.8) (UDP-N-
 DE acetylmutamoyl-L-alanine synthetase).
 GN Name=murC; OrderedlocusNames=XCC0726;
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xanthomonas.
 OK NCBI_TaxId=340;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33913 / NCPPB 528;
 RX MEDLINE=2202245; PubMed=12024217; DOI=10.1038/417459a;
 RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
 RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
 RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
 RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Clapina L.P.,
 RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
 RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
 RA Fomighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
 RA Katuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
 RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
 RA Martins E.C., Melandri J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
 RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
 RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
 RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
 RA Setubal J.C., Kitajima J.P.;
 RA "Comparison of the genomes of two Xanthomonas pathogens with differing
 RT host specificities";
 RL Nature 417:459-463 (2002).
 CC -1- FUNCTION: Cell wall formation.
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmutamate + L-alanine = ADP +
 CC phosphate + UDP-N-acetylmutamoyl-L-alanine.
 CC -1- PATHWAY: Peptidoglycan biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC -1- SIMILARITY: Belongs to the murCDP family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AE012171; AAM40041.1; -;
 CC DR HSSP; P45066; IG00.
 CC DR HAMAP; MF_00046; -; 1.
 CC DR InterPro; IPR004101; Mur_1lgase_C.
 CC DR InterPro; IPR000713; Mur_1lgase_N.
 CC DR InterPro; IPR005758; MurC.
 CC DR Pfam; PF01225; Mur_1lgase; 1.
 CC DR Pfam; PF02875; Mur_1lgase; 1.
 CC DR TIGRFAMs; TIGR01082; murC; 1.
 CC DR ATP-binding; Cell division; Cell wall; Complete proteome; Ligase;
 CC KW NP_BIND 122 128 ATP (potential).
 CC FT SEQUENCE 477 AA; 50524 MW; 787701B6E2310C89 CRC64;
 CC SQ
 Query Match 100.0%; Score 35; DB 1; Length 477;
 Best Local Similarity 100.0%; Pred. No. 67;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 DSGDLVR 7
 Db 7 DSGDLVR 13
 RESULT 4
 Q983X2 PRELIMINARY; PRT; 289 AA.
 ID Q983X2;
 AC 0983X2;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE M18130 protein.
 GN OrderedlocusNames=m18130;
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:15:35 ; Search time 69.1667 Seconds
(without alignments)
51.825 Million cell updates/sec

Title: US-10-646-919-35
Perfect score: 35
Sequence: 1 DSGDLVR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : uniprot_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	100.0	386	2 Q6DIE1	Q6DIE1 erwina car
2	35	100.0	477	1 MURC_XANAC	Q8PPA7 xanthomonas
3	35	100.0	477	1 MURC_XANCP	Q8PCJ9 xanthomonas
4	34	97.1	289	2 Q983X2	Q983X2 rhizobium 1
5	33	94.3	400	2 Q7RGE6	Q7RGE6 plasmodium
6	32	91.4	101	2 Q9QUU6	Q9QUU6 rattus sp.
7	32	91.4	160	2 Q983Q4	Q983Q4 rhizobium 1
8	32	91.4	167	2 Q6LBC7	Q6LBC7 oligotropha
9	32	91.4	313	2 Q8RSH0	Q8RSH0 rattus norv
10	32	91.4	452	2 Q7UW19	Q7UW19 rhodospirillum
11	32	91.4	595	1 PTN6_MOUSE	P29351 mus musculus
12	32	91.4	613	1 PTN6_RAT	P81718 rattus norv
13	32	91.4	656	2 Q8X438	Q8X438 escherichia
14	32	91.4	682	1 RHSE_ECOLI	P24211 escherichia
15	32	91.4	860	2 Q6D8U4	Q6D8U4 erwina car
16	32	91.4	1398	2 Q7AGM1	Q7AGM1 escherichia
17	32	91.4	1398	2 Q8XET5	Q8XET5 escherichia
18	32	91.4	1400	2 Q8X2F8	Q8X2F8 escherichia
19	32	91.4	1409	2 O52661	O52661 escherichia
20	32	91.4	1426	1 RHSD_ECOLI	P16919 escherichia
21	32	91.4	5289	2 Q88Z03	Q88Z03 lactobacillus
22	31	88.6	106	2 Q7DKC0	Q7DKC0 oryza sativ
23	31	88.6	184	2 Q26364	Q26364 methanobact
24	31	88.6	251	2 Q9L4W2	Q9L4W2 streptomyces
25	31	88.6	258	2 Q88JP8	Q88JP8 pseudomonas
26	31	88.6	263	2 Q72BF6	Q72BF6 desulfotomob
27	31	88.6	393	2 Q96RJ7	Q96RJ7 homo sapien
28	31	88.6	393	2 Q96RNO	Q96RNO homo sapien
29	31	88.6	394	2 Q85162	Q85162 pseudomonas
30	31	88.6	394	2 Q9HU25	Q9HU25 pseudomonas
31	31	88.6	400	2 Q8EG14	Q8EG14 shewanella

32	31	88.6	455	1 DD15_STRPU	Q17438 strongyloce
33	31	88.6	477	1 MURC_XYLFA	Q9P780 xyliella faa
34	31	88.6	477	1 MURC_XYLFT	Q87890 xyliella faa
35	31	88.6	595	1 PTN6_HUMAN	P29350 homo sapien
36	31	88.6	624	2 Q9UK67	Q9UK67 homo sapien
37	31	88.6	675	2 Q9RUC9	Q9RUC9 delnococtus
38	31	88.6	691	2 Q87019	Q87019 cryptococcus
39	31	88.6	710	2 Q7Q453	Q7Q453 anophelis g
40	31	88.6	879	2 Q7NHE2	Q7NHE2 gloeobacter
41	31	88.6	1006	2 P74686	P74686 synchocyst
42	31	88.6	4467	2 Q9J3E9	Q9J3E9 murine hepa
43	31	88.6	4470	2 Q66WN5	Q66WN5 murine hepa
44	31	88.6	7176	1 R1AB_CYMAS	P16342 m replicase
45	31	88.6	7178	2 Q66WN6	Q66WN6 murine hepa

ALIGNMENTS

RESULT 1					
ID	Q6DIE1	PRELIMINARY;	PRT;	386 AA.	
AC	Q6DIE1				
DT	25-OCT-2004 (TrEMBLrel. 28, Created)				
DT	25-OCT-2004 (TrEMBLrel. 28, Last sequence update)				
DE	Putative exported protein.				
GN	OrderedlocusNames=ECA3507;				
OS	Erwinia carotovora (subsp. atroseptica) (Pectobacterium atrosepticum).				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;				
OC	Enterobacteriaceae; Pectobacterium.				
OX	NCBI_Taxid=29471;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SCRI 1043 / ATCC BAA-672;				
RX	PubMed=15263089; DOI=10.1073/pnas.0402424101.				
RA	Bell K.S., Sebalth M., Pritchard L., Holden M.T.G., Hyman L.J.,				
RA	Holeva M.C., Thomson N.R., Bentley S.D., Churcher L.J.C., Mungall K.,				
RA	Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J.,				
RA	Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H.,				
RA	Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S.,				
RA	Salmond G.P.C., Birch P.R.U., Parkhill J., Toth I.K.,				
RT	"Genome sequence of the enterobacterial phytopathogen Erwinia				
RT	carotovora subsp. atroseptica and characterization of virulence				
RT	factors."				
RL	Proc. Natl. Acad. Sci. U.S.A. 101:11105-11110 (2004).				
DR	EMBL: BX950851; CAG76405.1; -.				
DR	InterPro: IPR011045; N20_reductase_N.				
KW	Complete proteome.				
SQ	SEQUENCE 386 AA; 41410 MW; B93562E770DDDE61 CRC64;				
Query Match					
Best Local Similarity 100.0%; Score 35; DB 2; Length 386;					
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
Qy	1 DSGDLVR 7				
Db	361 DSGDLVR 367				
RESULT 2					
MURC_XANAC	STANDARD;	PRT;	477 AA.		
ID	MURC_XANAC				
AC	Q8PPA7;				
DT	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	05-JUL-2004 (Rel. 44, Last annotation update)				
DE	UDP-N-acetylglucosamine--L-alanine ligase (EC 6.3.2.8) (UDP-N-				
DE	acetylglucosamoyl-L-alanine synthetase).				
GN	Name=MURC; OrderedlocusNames=XAC0780;				
OS	Xanthomonas axonopodis (pv. citri).				
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;				
OC	Xanthomonadaceae; Xanthomonas.				

A:Accession: A42031
A:Molecule type: mRNA
A:Residues: 1-595 <Y11>
A:Cross-references: GB:W68902; NID:g193807
A:Note: sequence extracted from NCBI backbone (NCBIN:76845, NCBI:76846)
R:Y1, T.; Cleveland, J.L.; Ihle, J.N.
Blood 78, 2222-2228, 1991
A:Title: Identification of novel protein tyrosine phosphatases of hematopoietic cells by
A:Reference number: A61180; MUID:92032882; PMID:1932742
A:Accession: F61180
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 342-451 <Y1A>
R:Schultz, L.D.; Schweitzer, P.A.; Rajan, T.V.; Yi, T.; Ihle, J.N.; Matthews, R.J.; Thomas
Cell 73, 1445-1454, 1993
A:Title: Mutations at the murine motheaten locus are within the hematopoietic cell prote
A:Reference number: 152816; MUID:93313972; PMID:8324828
A:Accession: 165741
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 334-353, 359-382 <SHU1>
A:Cross-references: GB:S63503; NID:g388449
A:Note: deletion mutation
A:Accession: 152816
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 350-358, 'EGQSPNPLPTPTSSSLVVOYHTQ', 359-366 <SHU2>
A:Cross-references: GB:S63763; NID:g388447
A:Note: insertion mutation
A:Accession: 165740
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 34-76, 'VPRPHIRAGCVTAAGCGRAD', <SHU3>
A:Cross-references: GB:S63764; NID:g388450
A:Note: frameshift mutation
R:Yeung, Y.G.; Berg, K.L.; Pixley, F.J.; Angeletti, R.H.; Stanley, E.R.
J. Biol. Chem. 267, 23447-23450, 1992
A:Title: Protein tyrosine phosphatase-1C is rapidly phosphorylated in tyrosine in macro
A:Reference number: A45143; MUID:93054686; PMID:1385421
A:Accession: A45143
A:Status: preliminary
A:Molecule type: protein
A:Residues: 137-139, 'X', 141-143, 'X', 145-151 <YE2>
A:Experimental source: BAC1.2P5 macrophage (NCBI:118519)
A:Note: sequence extracted from NCBI backbone (NCBI:118519)
A:Accession: B45143
A:Status: preliminary
A:Molecule type: protein
A:Residues: 54-56, 'X', 58, 'X', 60-61, 'X', 63-68 <YE2>
A:Experimental source: BAC1.2P5 macrophage
A:Note: sequence extracted from NCBI backbone (NCBI:118518)
C:Comment: This protein is found primarily in hematopoietic tissues.
C:Gene: me/HCPH; motheaten
C:Function:
A:Description: catalyzes hydrolysis of peptidyl-phosphotyrosine to release phosphate
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-phosph
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase
F:4-98/Domain: SH2 homology <SH2A>
F:110-211/Domain: SH2 homology <SH2B>
F:265-521/Domain: phosphatase catalytic domain #status predicted <P>
F:270-504/Domain: protein-tyrosine-phosphatase homology <PTP>
F:553/Active site: Cys (phosphocysteine intermediate) #status predicted
F:459/Binding site: substrate phosphate (Arg) #status predicted

Query Match 91.4%; Score 32; DB 1; Length 595;
Best Local Similarity 85.7%; Pred. No. 63;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
DB 401 DNGDLVR 407

RESULT 3
G85731
Rhs element associated protein Z2257 [imported] - Escherichia coli (strain O157:H7, sube
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: G85731
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouels, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: G85731
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-656 <STO>
A:Cross-references: UNIPROT:O8X438; GB:AE005174; NID:g12515236; PIDN:AA656315.1; GSPDB:G
A:Experimental source: strain O157:H7, substrain EDL933
C:Gene: Z2257

Query Match 91.4%; Score 32; DB 2; Length 656;
Best Local Similarity 85.7%; Pred. No. 71;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
DB 126 DNGDLVR 132

RESULT 4
C64898
rhaB protein - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: C64898; J50626; S16027
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: C64898
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-662 <BLAT>
A:Cross-references: UNIPROT:P24211; GB:AE000242; GB:U00096; NID:g1787720; PIDN:AACT4538
A:Experimental source: strain K-12, substrain MG1655
R:Sadosky, A.B.; Gray, J.A.; Hill, C.W.
Nucleic Acids Res. 19, 7177-7183, 1991
A:Title: The RhaD-E subfamily of Escherichia coli K-12.
A:Reference number: J50625; MUID:92115567; PMID:1766878
A:Accession: J50626
A:Molecule type: DNA
A:Residues: 9-355, 'X', 357-441, 'G', 443-682 <SA2>
A:Cross-references: GB:X60998; NID:g42735; PID:g42736
A:Note: the authors translated the codon CAG for residue 349 as Gln
C:Gene: rhaE
A:Map position: 32 min

Query Match 91.4%; Score 32; DB 2; Length 682;
Best Local Similarity 85.7%; Pred. No. 74;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
DB 147 DNGDLVR 153

RESULT 5
H90698
RhaD core protein with extension [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:32:16 ; Search time 14.5 Seconds
(without alignments)
46.449 Million cell updates/sec

Title: US-10-646-919-35

Perfect score: 35
Sequence: 1 DSGDLVR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	32	91.4	101	2 B38189	protein-tyrosine-P
2	32	91.4	595	1 A44390	protein-tyrosine-P
3	32	91.4	656	2 G85731	Rhs element associ
4	32	91.4	682	2 C64698	Rhs protein - Bsc
5	32	91.4	1398	2 H90698	Rhsd core protein
6	32	91.4	1398	2 B85549	hypothetical prote
7	32	91.4	1400	2 E90886	Rhs core protein
8	32	91.4	1426	2 H64780	rhsp protein precu
9	31	88.6	184	2 C69133	DNA-dependent RNA
10	31	88.6	394	2 D83002	multidrug resistanc
11	31	88.6	477	2 E82763	UDP-N-acetylmurama
12	31	88.6	595	1 S20825	protein-tyrosine-P
13	31	88.6	675	2 E75393	hypothetical prote
14	31	88.6	1006	2 S76892	hypothetical prote
15	30	85.7	315	2 T36417	probable formyltra
16	30	85.7	358	2 F87364	HLyd family secret
17	30	85.7	359	2 T35340	probable adenosine
18	30	85.7	361	2 G95410	probable ABC trans
19	30	85.7	426	2 H70390	conserved hyposhet
20	30	85.7	457	2 S44269	platelet-derived g
21	30	85.7	459	2 AC2116	hypothetical prote
22	30	85.7	469	2 S4196	rad22 protein - fl
23	30	85.7	510	2 T13882	cytochrome-c oxida
24	30	85.7	559	2 B95014	metallo-beta-lacta
25	30	85.7	560	2 B86661	hypothetical prote
26	30	85.7	561	2 H69077	DNA ligase - Metha
27	30	85.7	610	2 E97887	conserved hypochet
28	30	85.7	719	2 AC0867	2-acylglycerophosp
29	30	85.7	719	2 E91090	hypothetical prote

30	30	85.7	719	2 H85935	hypothetical prote
31	30	85.7	719	2 E65066	2-acylglycerophosp
32	30	85.7	788	2 S67595	hypothetical prote
33	30	85.7	920	2 C96831	hypothetical prote
34	30	85.7	926	2 G96563	probable coatomer
35	30	85.7	1353	2 T19691	hypothetical prote
36	30	85.7	2658	2 A86216	protein T23G18.2 l
37	29	82.9	96	2 B85781	hypothetical prote
38	29	82.9	96	2 F90932	hypothetical prote
39	29	82.9	96	2 D64931	hypothetical prote
40	29	82.9	159	2 T37098	hypothetical prote
41	29	82.9	204	2 A10620	probable bacteriop
42	29	82.9	228	2 G72631	hypothetical prote
43	29	82.9	241	2 H81278	phosphatidate cycl
44	29	82.9	262	1 C69097	3',5'-cyclic-nucle
45	29	82.9	275	2 A87487	phosphatidate cycl

ALIGNMENTS

RESULT 1

B38189

protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 6 - rat (fragment)

N/Alternate names: protein-tyrosine-phosphatase SH-FP1; PTPY142

C/Species: Rattus norvegicus (Norway rat)

C/Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 24-Apr-1998

C/Accession: B38189

R.Plutsky, J. Neel, B.G. Rosenberg, R.D. Proc. Natl. Acad. Sci. U.S.A. 89, 1123-1127, 1992

A/Title: Isolation of a src homology 2-containing tyrosine phosphatase.

A/Reference number: A38189; MUID:92141214; PMID:1736296

A/Accession: B38189

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: nucleic acid

A/Residues: 1-101 <PLU>

A/Experimental source: megakaryocyte

A/Note: sequence extracted from NCBI backbone (NCBIP:79621)

C/Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-phosphatase

C/Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase

F:1-101/Domain: protein-tyrosine-phosphatase homology (fragment) <PP>

F:94/Active site: Cys (phosphocysteine intermediate) #status predicted

F:100/Binding site: substrate phosphate (Arg) #status predicted

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: nucleic acid

A/Residues: 1-101 <PLU>

A/Experimental source: megakaryocyte

A/Note: sequence extracted from NCBI backbone (NCBIP:79621)

C/Superfamily: protein-tyrosine-phosphatase, nonreceptor type 6; protein-tyrosine-phosphatase

C/Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase

F:1-101/Domain: protein-tyrosine-phosphatase homology (fragment) <PP>

F:94/Active site: Cys (phosphocysteine intermediate) #status predicted

Query Match 91.4%; Score 32; DB 2; Length 101;

Best local similarity 85.7%; Pred. No. 8; Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DSGDLVR 7

Db 42 DNGDLVR 48

RESULT 2

A44390

protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 6 - mouse

N/Alternate names: hematopoietic cell phosphatase (HCP); protein-tyrosine-phosphatase 1C

C/Species: Mus musculus (house mouse)

C/Date: 03-May-1994 #sequence_revision 19-May-1994 #text_change 11-Jun-1999

C/Accession: A44390; A42031; F61180; I65741; I52816; I65740; A45143; B45143

Mol. Cell. Biol. 12, 2396-2405, 1992

A/Title: Characterization of hematopoietic intracellular protein tyrosine phosphatases;

and threonine-rich sequences.

A/Reference number: A44390; MUID:92236615; PMID:1373816

A/Accession: A44390

A/Molecule type: mRNA

A/Status: preliminary; not compared with conceptual translation

A/Molecule type: nucleic acid

A/Residues: 1-595 <MAT>

A/Experimental source: GB:M90389; NTD:G200550; PIDN:AAA40007.1; PID:G200551

A/Title: Protein tyrosine phosphatase containing SH2 domains: characterization, preferen

A/Reference number: A42031; MUID:92123209; PMID:1732748

```
RESULT 2
US-10-425-115-354248
; Sequence 354248, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5322)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: MRT4577_86246C.1.pep
US-10-425-115-354248

Query Match
Best Local Similarity 91.4%; Score 32; DB 16; Length 112;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
Db 73 DSGDLVR 79

RESULT 3
US-10-366-547-97
; Sequence 97, Application US/10366547
; Publication No. US20030215899A1
; GENERAL INFORMATION:
; APPLICANT: Meng, Tzu-Ching
; APPLICANT: Tonks, Nicholas K.
; APPLICANT: Cool, Deborah E.
; TITLE OF INVENTION: REVERSIBLE OXIDATION OF PROTEIN TYROSINE
; FILE REFERENCE: 200125,439
; CURRENT APPLICATION NUMBER: US/10/366,547
; CURRENT FILING DATE: 2003-02-12
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 97
; LENGTH: 613
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-10-366-547-97

Query Match
Best Local Similarity 91.4%; Score 32; DB 15; Length 613;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
Db 403 DSGDLVR 409

RESULT 4
US-10-282-122A-56547
; Sequence 56547, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Hasebeck, Robert
; APPLICANT: Ohlsen, Karl
```

```
APPLICANT: Zyskind, Judith
APPLICANT: Mail, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Foreyth, R.
APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56547
; LENGTH: 682
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-56547

Query Match
Best Local Similarity 91.4%; Score 32; DB 15; Length 682;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
Db 147 DSGDLVR 153

RESULT 5
US-09-912-020-340
; Sequence 340, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Trawick, John
; APPLICANT: Foreyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001DIV1
; CURRENT APPLICATION NUMBER: US/09/912,020
; CURRENT FILING DATE: 2001-07-23
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
```


Sequence 13, Application US/09144925
Patent No. 5951979
GENERAL INFORMATION:
APPLICANT: Tonks, Nicholas
TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
TITLE OF INVENTION: TYROSINE PHOSPHATASES
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESS: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02421-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/144,925
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/685,992
FILING DATE: JULY 25, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: CSH96-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: 781-861-6240
TELEFAX: 781-861-9540
TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 263 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-144-925-13

Query Match 88.6%; Score 31; DB 2; Length 263;
Best Local Similarity 71.4%; Pred. No. 57;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
DB 141 DNGDLIR 147

RESULT 10
US-09-252-991A-23319
Sequence 23319, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 23319
LENGTH: 495
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-23319
Query Match 88.6%; Score 31; DB 4; Length 495;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
DB 174 DDGDLVR 180

RESULT 11
US-08-202-389-6
Sequence 6, Application US/08202389
Patent No. 553636
GENERAL INFORMATION:
APPLICANT: Freeman Jr., Robert M.
APPLICANT: Plutsky, Jorge
APPLICANT: Neel, Benjamin G.
APPLICANT: Rosenberg, Robert D.
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE
TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,389
FILING DATE: 28-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/983,926
FILING DATE: 01-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/829,141
FILING DATE: 31-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/721,112
FILING DATE: 26-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BI992-05WA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 595 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-202-389-6

Query Match 88.6%; Score 31; DB 1; Length 595;
Best Local Similarity 71.4%; Pred. No. 1.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
DB 401 DNGDLIR 407

RESULT 12

TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
FILE REFERENCE: TSRI 645.2
CURRENT APPLICATION NUMBER: US/09/494,190
CURRENT FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: EP/99/07742
PRIOR FILING DATE: 1999-10-14
PRIOR APPLICATION NUMBER: US 09/173,941
PRIOR FILING DATE: 1998-10-16
NUMBER OF SEQ ID NOS: 126
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 36
LENGTH: 7
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: phage display
OTHER INFORMATION: selected and mutagenized
US-09-494-190-36

Query Match 88.6%; Score 31; DB 4; Length 7;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
DB 1 DSGELVR 7

RESULT 7
US-08-202-389-18
Sequence 18, Application US/08202389
Patent No. 5536636
GENERAL INFORMATION:
APPLICANT: Freeman Jr., Robert M.
APPLICANT: Plutsky, Jorge
APPLICANT: Neel, Benjamin G.
APPLICANT: Roseberry, Robert D.
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE
TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,389
FILING DATE: 28-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/983,926
FILING DATE: 01-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/829,141
FILING DATE: 31-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/721,112
FILING DATE: 26-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BI92-05MA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:
LENGTH: 97 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-202-389-18

Query Match 88.6%; Score 31; DB 1; Length 97;
Best Local Similarity 71.4%; Pred. No. 19;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
DB 42 DNGDLVR 48

RESULT 8
US-08-685-992-13
Sequence 13, Application US/08685992
Patent No. 5912138
GENERAL INFORMATION:
APPLICANT: Tonks, Nicholas
APPLICANT: Flint, Andrew J.
TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
TITLE OF INVENTION: TYROSINE PHOSPHATASES
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,992
FILING DATE: 25-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: CSHL96-03
TELECOMMUNICATION INFORMATION:
TELEPHONE: 781-861-6240
TELEFAX: 781-861-9540
TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 263 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-685-992-13

Query Match 88.6%; Score 31; DB 2; Length 263;
Best Local Similarity 71.4%; Pred. No. 57;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
DB 141 DNGDLVR 147

RESULT 9
US-09-144-925-13

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:phage display
OTHER INFORMATION: selected and mutagenized
US-09-494-190-35

Query Match 100.0%; Score 35; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
DB 1 DSGDLVR 7

RESULT 3
US-08-202-389-2
Sequence 2, Application US/08202389
Patent No. 5536636
GENERAL INFORMATION:
APPLICANT: Freeman Jr., Robert M.
APPLICANT: Plutzky, Jorge
APPLICANT: Neel, Benjamin G.
APPLICANT: Rosenberg, Robert D.
TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE
TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/202,389
FILING DATE: 28-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/983,926
FILING DATE: 01-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/829,141
FILING DATE: 31-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/721,112
FILING DATE: 26-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: BIH92-05WA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 513 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-202-389-2

Query Match 91.4%; Score 32; DB 1; Length 513;
Best Local Similarity 85.7%; Pred. No. 72;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
DB 319 DNGDLVR 325

RESULT 4
US-09-492-709A-340
Sequence 340, Application US/09492709A
Patent No. 6720139
GENERAL INFORMATION:
APPLICANT: Zyskind, Judith
APPLICANT: Ohlsen, Kari L.
APPLICANT: Trawick, John
APPLICANT: Forsyth, R. Allyn
APPLICANT: Froelich, Jamie M.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
TITLE OF INVENTION: ESCHERICHIA COLI
FILE REFERENCE: ELITRA.001A
CURRENT APPLICATION NUMBER: US/09/492,709A
CURRENT FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 485
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 340
LENGTH: 1426
TYPE: PRT
ORGANISM: E. COLI
US-09-492-709A-340

Query Match 91.4%; Score 32; DB 4; Length 1426;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
DB 872 DNGDLVR 878

RESULT 5
US-09-173-941-36
Sequence 36, Application US/09173941
Patent No. 6140081
GENERAL INFORMATION:
APPLICANT: BARBAS, Carlos F.
TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
FILE REFERENCE: NOV0081S
CURRENT APPLICATION NUMBER: US/09/173,941
CURRENT FILING DATE: 1998-10-16
NUMBER OF SEQ ID NOS: 120
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 36
LENGTH: 7
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: nucleotide
US-09-173-941-36

Query Match 88.6%; Score 31; DB 3; Length 7;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|:|||||
DB 1 DSGDLVR 7

RESULT 6
US-09-494-190-36
Sequence 36, Application US/09494190
Patent No. 6610512
GENERAL INFORMATION:
APPLICANT: BARBAS, Carlos F.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:17:05 ; Search time 21.6667 Seconds

(without alignments)
24.117 Million cell updates/sec

Title: US-10-646-919-35

Perfect score: 35

Sequence: 1 DSGDLVR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	7	3 US-09-173-941-35	Sequence 35, App1
2	35	100.0	7	4 US-09-494-190-35	Sequence 35, App1
3	32	91.4	513	1 US-08-202-389-2	Sequence 2, App1
4	32	91.4	1426	4 US-09-492-709A-340	Sequence 340, App1
5	31	88.6	7	3 US-09-173-941-36	Sequence 36, App1
6	31	88.6	7	4 US-09-494-190-36	Sequence 36, App1
7	31	88.6	97	1 US-08-202-389-18	Sequence 18, App1
8	31	88.6	263	2 US-08-685-892-13	Sequence 13, App1
9	31	88.6	263	2 US-09-144-925-13	Sequence 13, App1
10	31	88.6	495	4 US-09-252-991A-23119	Sequence 23119, A
11	31	88.6	595	1 US-08-202-389-6	Sequence 6, App1
12	31	88.6	621	4 US-09-949-016-8360	Sequence 8360, App1
13	31	88.6	631	1 US-08-202-389-8	Sequence 8, App1
14	31	88.6	843	4 US-09-417-197-117	Sequence 117, App1
15	31	88.6	853	4 US-09-417-197-119	Sequence 119, App1
16	30	85.7	523	4 US-09-555-889A-2	Sequence 2, App1
17	30	85.7	559	4 US-09-583-110-4863	Sequence 4863, App1
18	30	85.7	612	4 US-09-107-433-2804	Sequence 2804, App1
19	29	82.9	7	3 US-09-173-941-50	Sequence 50, App1
20	29	82.9	7	4 US-09-494-190-50	Sequence 50, App1
21	29	82.9	162	4 US-09-902-540-9741	Sequence 9741, App1
22	29	82.9	312	4 US-09-252-991A-19374	Sequence 19374, A
23	29	82.9	332	4 US-09-902-540-9978	Sequence 9978, App1
24	29	82.9	349	4 US-09-252-991A-27933	Sequence 27933, A
25	29	82.9	401	3 US-08-861-774E-88	Sequence 88, App1
26	29	82.9	401	4 US-09-489-039A-12491	Sequence 12491, A
27	29	82.9	582	2 US-08-403-852D-20	Sequence 20, App1

28	29	82.9	582	3 US-08-510-646B-21	Sequence 21, App1
29	29	82.9	582	3 US-09-231-818-20	Sequence 20, App1
30	29	82.9	582	4 US-09-635-359B-20	Sequence 20, App1
31	29	82.9	794	4 US-09-134-000C-5518	Sequence 5518, App1
32	29	82.9	817	4 US-09-489-039A-10407	Sequence 10407, A
33	29	82.9	927	4 US-09-198-452A-472	Sequence 472, App1
34	29	82.9	937	4 US-09-438-185A-449	Sequence 449, App1
35	29	82.9	964	4 US-09-902-540-14068	Sequence 14068, A
36	29	82.9	1128	4 US-09-252-991A-31032	Sequence 31032, A
37	29	82.9	1180	4 US-09-252-991A-25722	Sequence 25722, A
38	29	82.9	2323	4 US-09-710-279-760	Sequence 760, App1
39	29	82.9	2404	3 US-09-134-001C-3464	Sequence 3464, App1
40	28	80.0	71	4 US-09-902-540-13751	Sequence 13751, A
41	28	80.0	169	4 US-09-489-039A-11124	Sequence 11124, A
42	28	80.0	171	4 US-09-270-767-34902	Sequence 34902, A
43	28	80.0	171	4 US-09-270-767-50119	Sequence 50119, A
44	28	80.0	235	4 US-09-252-991A-29626	Sequence 29626, A
45	28	80.0	244	4 US-09-248-796A-18903	Sequence 18903, A

ALIGNMENTS

```
RESULT 1
US-09-173-941-35
; Sequence 35, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV0081S
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 35
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
; OTHER INFORMATION: codon binding sequence
US-09-173-941-35

Query Match          100.0%; Score 35; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSGDLVR 7
       |||||
Db      1 DSGDLVR 7

RESULT 2
US-09-494-190-35
; Sequence 35, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 35
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

XX
PS Disclosure; Fig 1; 48pp; English.

CC The invention relates to zinc finger nucleotide-binding proteins which
CC comprise 2-12, preferably 2-6, operatively linked motifs selected from
CC sequences AAB02860-B02875. Sequences AAB02860-B02875 represent the alpha
CC helical regions of zinc finger domains which specifically bind to target
CC nucleotide triplets of the sequence 5'-GNN-3'. Such regions may be linked
CC by the peptide linker TGERP (AAB02970). The Cys2-His2 zinc finger motif
CC is the most frequently utilized nucleic acid binding motif in eukaryotes,
CC and constitutes a beta-beta-alpha fold. Nucleic acid recognition is
CC achieved through specific contacts from side chains of amino acid
CC residues in the alpha helix. Each zinc finger can recognise a sub-site of
CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains
CC allows the recognition of extended contiguous asymmetric DNA sequences.
CC For example, a synthetic polypeptide protein containing six zinc finger
CC domains can recognise an 18 bp sequence, and such proteins are
CC potentially highly gene-specific. The novel nucleotide-binding zinc
CC finger proteins may therefore be used in the development of artificial
CC gene-specific transcriptional regulators. Such transcriptional switches
CC may be used to regulate the expression of oncogenes such as erbB-2,
CC overexpression of which is involved in malignant transformation. The
CC proteins are therefore useful in the treatment of cancers, and may also
CC be used to activate genes involved in fighting diseases, and to treat
CC viral infections by inhibiting the synthesis of viral gene products. They
CC may be used in DNA-based diagnostic applications. The proteins may also
CC be used in producing functional gene knockout or activation in
CC heterozygous transgenic animals. Proteins of the invention can
CC discriminate between sequences which have a single base difference. This
CC is manifested in a >100-fold decrease in affinity for the variant
CC sequence. Gene activation and repression can be achieved by targeting
CC within the gene transcript, suggesting that information obtained from
CC expressed sequence tags may be sufficient for the construction of gene
CC switches. Sequences AAB02876-B02869 represent zinc finger alpha helix
CC phage library peptides disclosed in the invention
XX
SQ Sequence 7 AA;

Query Match 88.6%; Score 31; DB 3; Length 7;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|||:||||
DB 1 DSGELVR 7

RESULT 15

ADFE1985
ID ADF61985 standard; peptide; 7 AA.

AC ADF61985;

DT 12-FEB-2004 (first entry)

XX zinc finger binding region #36.

XX zinc finger-nucleotide binding polypeptide; expression regulation;

XX zinc finger binding region.

OS Synthetic.

XX US6610512-B1.

XX 26-AUG-2003.

XX 28-JAN-2000; 2000US-00494190.

XX 16-OCT-1998; 98US-00173941.

XX 14-OCT-1999; 99WO-EP007742.

XX (SCRI) SCRIPPS RES INST.

PI Barbas CF;
XX
XX WPI, 2003-800134/75.

PT Regulating expression of nucleotide sequence that contains sequence 5'-
PT (GNN)n-3', comprising exposing nucleotide sequence to composition
PT comprising isolated and purified zinc finger-nucleotide binding
PT polypeptide.

PS Disclosure; SEQ ID NO 36; 46pp; English.

XX
XX The invention relates to an expression of nucleotide sequence that
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by
CC exposing nucleotide sequence to composition comprising isolated and
CC purified zinc finger-nucleotide binding polypeptide that contains a
CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Leu-Val-
CC Arg. The invention is used in the regulation of the expression of the
CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present
CC sequence represents the amino acid sequence of a zinc finger binding
CC region.

XX
SQ Sequence 7 AA;

Query Match 88.6%; Score 31; DB 7; Length 7;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
|||:||||
DB 1 DSGELVR 7

Search completed: July 20, 2005, 13:45:15
Job time : 83.5 secs

XX AA65809 to AA65889 and AA66058 to AA66138 represent nucleotide
CC sequences derived from *Escherichia coli* which inhibit *E. coli*
CC proliferation. AA65890 to AA66055 and AA65886 to AA66040 represent
CC nucleotide and protein sequences associated with *E. coli* proliferation.
CC AA66056 and AA66057 represent primers used for sequencing *E. coli*
CC proliferation inhibiting nucleotide inserts in an example from the
CC present invention. Methods from the present invention can be used to
CC identify a proliferation-regulated gene in a microorganism, by contacting
CC a microorganism with a proliferation-regulated gene activity inhibitory
CC nucleic acid identified in another organism, and determining if
CC inhibition occurs in the second microorganism. The nucleic acid sequences
CC identified as being required for bacterial growth and proliferation, can
CC be used for antisense therapy for killing bacteria

SQ Sequence 1426 AA;

Query Match 91.4%; Score 32; DB 3; Length 1426;
Best Local Similarity 85.7%; Pred. No. 8.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 DSGDVLVR 7
|:|||||
Db 872 DNGDVLVR 878

RESULT 13
ABU14693
ID ABU14693 standard; protein; 1426 AA.
XX ABU14693;
AC
XX 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #220.
XX
KM Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS *Escherichia coli*.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX PF
XX 21-MAR-2001; 2001US-00815243.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923F.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699F.
XX
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haeelbeck R, Ohlsen KL, Zyskind JW,
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI: 2003-029926/02.
XX N-PSDB; ACA18563.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 42617; 1766bp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the

CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-regulated gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 1426 AA;

Query Match 91.4%; Score 32; DB 6; Length 1426;
Best Local Similarity 85.7%; Pred. No. 8.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 DSGDVLVR 7
|:|||||
Db 872 DNGDVLVR 878

RESULT 14
AAB02895
ID AAB02895 standard; peptide; 7 AA.
XX AAB02895;
AC
XX 18-SEP-2000 (first entry)
XX
DE Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:36.
XX
XX
XX Zinc finger domain; alpha helix; nucleotide binding; DNA binding;
XX polydactyl protein; asymmetric target recognition;
XX gene specific transcriptional regulator; gene activator; gene repressor;
XX transcriptional switch; oncogene; erbB-2; cancer; tumour; gene therapy;
XX transgenic animal; antiviral; anticancer; diagnosis.
XX
OS Synthetic.
XX
XX WO200023464-A2.
XX PN
XX 27-APR-2000.
XX PD
XX 14-OCT-1999; 99WO-EP007742.
XX PF
XX 16-OCT-1998; 98US-00173941.
XX PR
XX (NOVS) NOVARTIS AG.
XX PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX PA (SCRI) SCRIPPS RES INST.
XX
XX Barbas CF;
XX PI
XX WPI: 2000-339648/29.
XX DR
XX Novel isolated and purified zinc finger nucleotide-binding proteins with
PT specificity for GNN triplet sequences, useful in gene therapy and for
PT regulating gene function.

KM food supplement; medical imaging; diagnostic; genetic disorder.
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US008631.
 PF
 XX 31-MAR-2000; 2000US-00540217.
 PR
 XX 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 PI WPI; 2001-639362/73.
 DR N-PSDB; AAS79984.
 DR
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 46156; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 CC
 XX
 SQ Sequence 973 AA;
 Query Match 91.4%; Score 32; DB 4; Length 973;
 Best Local Similarity 85.7%; Pred. No. 5.6e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSGDLVR 7
 DB 866 DNGDLVR 872
 RESULT 11
 ADC01365
 ID ADC01365 standard; protein; 1400 AA.
 XX
 AC ADC01365;
 XX
 DT 04-DEC-2003 (first entry)
 DE Enterohemorrhagic E. coli O157:H7-specific protein SEQ ID NO: 1410.
 XX
 KM enterohaemorrhagic; anti-bacterial.
 XX
 OS Escherichia coli; O157:H7.

XX
 PN JP2002355074-A.
 XX
 PD 105DEC-2002.
 XX
 XX 24-JAN-2002; 2002JP-00015959.
 PF
 XX 24-JAN-2001; 2001JP-00112010.
 PR
 XX (UYTS-) UNIV TSUKUBA.
 PA WPI; 2003-451640/43.
 DR
 XX
 PT Enterohemorrhagic Escherichia coli O157:H7-specific nucleic acid molecule
 PT and a polypeptide and its use, a polypeptide, a vector and a host cell.
 XX
 PS Claim 3; SEQ ID NO 1410; 2067pp; Japanese.
 XX
 CC The invention relates to a novel enterohaemorrhagic Escherichia coli
 CC O157:H7-specific nucleic acid molecule. A polynucleotide of the invention
 CC has anti-bacterial activity. The polypeptide can be used in detection
 CC and/or treatment of O157:H7 infection. The nucleotide sequence of the
 CC genome of Enterohaemorrhagic E. coli O157:H7 was determined. The present
 CC sequence represents an E. coli O157:H7-specific polypeptide of the
 CC invention.
 CC
 XX
 SQ Sequence 1400 AA;
 Query Match 91.4%; Score 32; DB 7; Length 1400;
 Best Local Similarity 85.7%; Pred. No. 8.5e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 DSGDLVR 7
 DB 870 DNGDLVR 876
 RESULT 12
 AAB15983
 ID AAB15983 standard; protein; 1426 AA.
 XX
 AC AAB15983;
 XX
 DT 05-OCT-2000 (first entry)
 DE E. coli proliferation associated protein sequence SEQ ID NO:340.
 XX
 XX Escherichia coli; E. coli; proliferation; inhibition; screening;
 KM antimicrobial; bacterial growth; antisense therapy; antibacterial.
 XX
 OS Escherichia coli.
 OS
 PN WO2000044906-A2.
 PD 03-AUG-2000.
 PF 27-JAN-2000; 2000WO-US002200.
 XX
 XX 27-JAN-1999; 99US-0117405P.
 PR
 XX (ELITR-) ELITRA PHARM INC.
 PA
 XX Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM;
 PI Carr GJ, Yamamoto RT, Xu HH;
 XX
 DR WPI; 2000-514822/46.
 DR N-PSDB; AAB65988.
 XX
 PT Novel polynucleotides and polypeptides associated with microorganism
 PT proliferation, used to identify inhibitors of bacterial growth and
 PT proliferation, for use in antisense therapy.
 XX
 XX Claim 11; Page 255-256; 316pp; English.

XX	21-MAR-2001; 2001US-00815242.	
PR	06-SEP-2001; 2001US-00948993.	
PR	25-OCT-2001; 2001US-0342923P.	
PR	08-FEB-2002; 2002US-00072851.	
PR	06-MAR-2002; 2002US-0362699P.	
XX		
PA	(ELIT-) ELITRA PHARM INC.	
PI	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW,	
PI	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,	
DR	WPI: 2003-029926/02.	
XX	N-PSDB; ACA32493.	
PT	New antisense nucleic acids, useful for identifying proteins or screening	
PT	for homologous nucleic acids required for cellular proliferation to	
PT	isolate candidate molecules for rational drug discovery programs.	
XX		
PS	Claim 25; SEQ ID NO 56547; 1766pp; English.	
XX		
CC	The invention relates to an isolated nucleic acid comprising any one of	
CC	the 6213 antisense sequences given in the specification where expression	
CC	of the nucleic acid inhibits proliferation of a cell. Also included are:	
CC	(1) a vector comprising a promoter operably linked to the nucleic acid	
CC	encoding a polypeptide whose expression is inhibited by the antisense	
CC	nucleic acid; (2) a host cell containing the vector; (3) an isolated	
CC	polypeptide or its fragment whose expression is inhibited by the	
CC	antisense nucleic acid; (4) an antibody capable of specifically binding	
CC	the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular	
CC	proliferation, or the activity of a gene in an operon required for	
CC	proliferation; (7) identifying a compound that influences the activity of	
CC	the gene product or that has an activity against a biological pathway	
CC	required for proliferation, or that inhibits cellular proliferation; (8)	
CC	identifying a gene required for cellular proliferation or the biological	
CC	pathway in which a proliferation-required gene or its gene product lies	
CC	or a gene on which the test compound that inhibits proliferation of an	
CC	organism acts; (9) manufacturing an antibiotic; (10) profiling a	
CC	compound's activity; (11) a culture comprising strains in which the gene	
CC	product is overexpressed or underexpressed; (12) determining the extent	
CC	to which each of the strains is present in a culture or collection of	
CC	strains; or (13) identifying the target of a compound that inhibits the	
CC	proliferation of an organism. The antisense nucleic acids are useful for	
CC	identifying proteins or screening for homologous nucleic acids required	
CC	for cellular proliferation to isolate candidate molecules for rational	
CC	drug discovery programs, or for screening homologous nucleic acids	
CC	required for proliferation in cells other than <i>S. aureus</i> , <i>S. typhimurium</i> ,	
CC	<i>K. pneumoniae</i> or <i>P. aeruginosa</i> . The present sequence is encoded by one of	
CC	the target prokaryotic essential genes. Note: The sequence data for this	
CC	patent did not form part of the printed specification, but was obtained	
CC	in electronic format directly from WIPO at	
CC	ftp.wipo.int/pub/published_pdc_sequences	
XX		
QQ	Sequence 682 AA;	
XX		
Query Match	91.4%; Score 324; DB 6; Length 682;	
Best Local Similarity	85.7%; Pred. No. 3,8e+02;	
Matches	6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;	
OY	1 DSGDIV 7	
DB	147 DNGDLVR 153	
RESULT 9		
ABG09197		
AC	ABG09197 standard; protein; 833 AA.	
XX		
XX	ABG09197;	
DT	13-FEB-2002 (first entry)	
XX		
DE	Novel human diagnostic protein #9198.	

XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
XX	Homo sapiens.
OS	
XX	WO200175067-A2.
PN	
XX	11-OCT-2001.
PD	
XX	
XX	30-MAR-2001; 2001WO-US008631.
PF	
XX	
XX	31-MAR-2000; 2000US-00540217.
FR	
XX	23-AUG-2000; 2000US-00649167.
PR	
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Dmanac RT, Liu C, Tang YT;
XX	
DR	WPI; 2001-639362/73.
DR	N-PSDB; AAS73384.
XX	
XX	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity.
XX	
PS	Claim 20; SEQ ID NO 39556; 103pp; English.
XX	
XX	The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC	sequences. (I) is useful as hybridization probes, polymerase chain
CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC	and in recombinant production of (II). The polynucleotides are also used
CC	in diagnostics as expressed sequence tags for identifying expressed
CC	genes. (II) is useful in gene therapy techniques to restore normal
CC	activity of (II) or to treat disease states involving (II). (II) is
CC	useful for generating antibodies against it, detecting or quantitating a
CC	polypeptide in tissue, as molecular weight markers and as a food
CC	supplement. (II) and its binding partners are useful in medical imaging
CC	of sites expressing (II). (I) and (II) are useful for treating disorders
CC	involving aberrant protein expression or biological activity. The
CC	polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC	amino acid sequences of the invention. Note: The sequence data for this
CC	patent did not appear in the printed specification, but was obtained in
CC	electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	
XX	
SO	Sequence 833 AA;
XX	
XX	
XX	Query Match 91.4%; Score 32; DB 4; Length 833;
XX	Best Local Similarity 85.7%; Pred. No. 4.7e+02;
XX	Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY	1 DSGDLVR 7
XX	:
DB	749 DNGDLVR 755
XX	
XX	
XX	RESULT 10
XX	ABG15797
ID	ABG15797 standard; protein; 973 AA.
XX	
AC	ABG15797;
XX	
DT	18-FEB-2002 (first entry)
XX	
XX	Novel human diagnostic protein #15788.
XX	
XX	
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;


```

RESULT 4
AAR56994
XX AAR56994 standard; protein; 161 AA.
AC
XX
XX AAR56994;
AC
XX
XX 25-MAR-2003 (revised)
DT 28-FEB-1995 (first entry)
XX
XX MPTP-H3/180.
DE
XX Protein tyrosine phosphatase; PTP; primer; PCR; amplification;
KW polymerase chain reaction; restriction analysis; transcription;
KM multigene; reverse transcriptase.
XX
XX Mus musculus.
OS
XX DE4242638-A1.
PN
XX 23-JUN-1994.
PD
XX
XX 17-DEC-1992; 92DE-04242638.
PF
XX
XX 17-DEC-1992; 92DE-04242638.
PR
XX
XX (BOEH/) BOEHM T.
PA
XX
XX Boehm T;
PI
XX
XX MPI; 1994-235494/29.
DR
XX
XX Analysing transcription patterns of gene and multi-gene families - by
PT incomplete PCR with only one labelled primer and restriction analysis of
PT amplification products, also new protein tyrosine phosphate and DNA
PT encoding it.
XX
XX Example 4; Page 14; 22pp; German.
XX
XX Analysis of transcription patterns of gene or multigene families
CC comprises (1) isolating RNA and subjecting it to reverse transcriptase;
CC (2) amplifying the cDNA using one labelled and one unlabelled primer, or
CC primer sets; (3) stopping amplification before conversion of starting
CC materials is complete; and (4) subjecting the DNA produced to restriction
CC analysis. The primer given in AAO57682 is used in the amplification of
CC MPTP-H3/180 (given in AAR56994). (Updated on 25-MAR-2003 to correct PN
CC field.)
XX
XX
XX Sequence 161 AA;
SQ
XX
XX
XX Query Match 91.4%; Score 32; DB 2; Length 161;
Best Local Similarity 85.7%; Pred. No. 73;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSGDLVR 7
|:|||||
Db 104 DNGDLVR 110

RESULT 5
ABG09201
ID ABG09201 standard; protein; 301 AA.
XX
XX
XX ABG09201;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #9192.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX

```

```

OS Homo sapiens.
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US008631.
PF
XX
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Dmanac RT, Liu C, Tang YT;
PI
XX
XX MPI; 2001-639362/73.
DR N-FSDB; AAS73386.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 39560; 103bp; English.
PS
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX
XX Sequence 301 AA;
SQ
XX
XX
XX Query Match 91.4%; Score 32; DB 4; Length 301;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSGDLVR 7
|:|||||
Db 95 DNGDLVR 101

RESULT 6
AAR99314
ID AAR99314 standard; protein; 513 AA.
XX
XX
XX AAR99314;
AC
XX
XX 25-MAR-2003 (revised)
DT 13-NOV-1996 (first entry)
XX
XX Rat MLPTP (protein tyrosine phosphatase-1) clone.
DE
XX
XX PTP; protein tyrosine phosphatase; SH2; Src homology region 2;
KW chromosome 12p; abnormality; mutation; detection; probe; neoplasia;
KW cancer; leukaemia; diagnosis; megakaryocyte regulation.
XX
XX

```

CC achieved through specific contacts from side chains of amino acid
CC residues in the alpha helix. Each zinc finger can recognise a subunit of
CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains
CC allows the recognition of extended contiguous asymmetric DNA sequences.
CC For example, a synthetic polypeptide protein containing six zinc finger
CC domains can recognise an 18 bp sequence, and such proteins are
CC potentially highly gene-specific. The novel nucleotide-binding zinc
CC finger proteins may therefore be used in the development of artificial
CC gene-specific transcriptional regulators. Such transcriptional switches
CC may be used to regulate the expression of oncogenes such as erbB-2,
CC overexpression of which is involved in malignant transformation. The
CC proteins are therefore useful in the treatment of cancers, and may also
CC be used to activate genes involved in fighting diseases, and to treat
CC viral infections by inhibiting the synthesis of viral gene products. They
CC may be used in DNA-based diagnostic applications. The proteins may also
CC be used in producing functional gene knockout or activation in
CC heterozygous transgenic animals. Proteins of the invention can
CC discriminate between sequences which have a single base difference. This
CC is manifested in a >100-fold decrease in affinity for the variant
CC sequence. Gene activation and repression can be achieved by targeting
CC within the gene transcript, suggesting that information obtained from
CC expressed sequence tags may be sufficient for the construction of gene
CC switches. Sequences AAB02876-B02869 represent zinc finger alpha helix
CC phage library peptides disclosed in the invention
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 35; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
| | | | |
Db 1 DSGDLVR 7

RESULT 2
ADP61984
ID ADP61984 standard; peptide; 7 AA.

AC ADP61984;

DT 12-FEB-2004 (first entry)

XX zinc finger binding region #35.

XX zinc finger-nucleotide binding polypeptide; expression regulation;

XX zinc finger binding region.

OS Synthetic.

PN US6610512-B1.

PD 26-AUG-2003.

PF 28-JUN-2000; 2000US-00494190.

PR 16-OCT-1998; 98US-00173941.

PR 14-OCT-1999; 99WO-EP007742.

PA (SCRI) SCRIPPS RES INST.

PI Barbas CF;

XX MPI; 2003-800134/75.

PT Regulating expression of nucleotide sequence that contains sequence 5'-
PT (GNN)n-3', comprising exposing nucleotide sequence to composition
PT comprising isolated and purified zinc finger-nucleotide binding
PT polypeptide.

XX Disclosure; SEQ ID NO 35; 46pp; English.

CC The invention relates to an expression of nucleotide sequence that
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by
CC exposing nucleotide sequence to composition comprising isolated and
CC purified zinc finger-nucleotide binding polypeptide that contains a
CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Leu-Val-
CC Arg. The invention is used in the regulation of the expression of the
CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present
CC sequence represents the amino acid sequence of a zinc finger binding
CC region.
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 35; DB 7; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
| | | | |
Db 1 DSGDLVR 7

RESULT 3
ADJ98426
ID ADJ98426 standard; peptide; 7 AA.

AC ADJ98426;

DT 06-MAY-2004 (first entry)

XX zinc finger DNA binding peptide #97.

XX library: multimeric DNA binding polypeptide;
XX zinc finger DNA binding peptide; gene expression silencing;
XX gene expression enhancement.
XX

OS Unidentified.

PN WO2003066828-A2.

PD 14-AUG-2003.

PF 07-FEB-2003; 2003WO-US003705.

PR 07-FEB-2002; 2002US-0354981P.

XX (SCRI) SCRIPPS RES INST.

XX Barbas CF, Blancafort P;

PI MPI; 2003-731499/69.

XX New zinc finger library of multimeric DNA binding polypeptides, useful
XX PT for sterically occluding the binding site of a natural transcription
XX PT factor, and enhancing or silencing target gene expression.
XX

PS Disclosure; Fig 20; 64pp; English.

XX The invention comprises a library of multimeric DNA binding polypeptides
XX CC (e.g. zinc finger DNA binding polypeptides). The zinc finger binding
XX CC polypeptides of the invention are useful for sterically occluding the
XX CC binding site of a natural transcription factor and enhancing or silencing
XX CC target gene expression. The present amino acid sequence represents a zinc
XX CC finger DNA binding peptide of the invention.
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 35; DB 7; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSGDLVR 7
| | | | |
Db 1 DSGDLVR 7

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:14:30 ; Search time 82.5 Seconds
(without alignments)
32.816 Million cell updates/sec

Title: US-10-646-919-35

Perfect score: 35

Sequence: 1 DSGDLVR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: A_Geneseq_16pcc04:*
2: geneseqp19808:*
3: geneseqp19908:*
4: geneseqp20008:*
5: geneseqp20018:*
6: geneseqp20028:*
7: geneseqp20038:*
8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	7	3 AAB02894	Aab02894 Nucleotid
2	35	100.0	7	7 ADF61984	Adf61984 Zinc fing
3	35	100.0	7	7 ADJ98426	Adj98426 Zinc fing
4	32	91.4	161	2 AAR56994	Aar56994 WPTP-H3/L
5	32	91.4	301	4 ABG09201	Abg09201 Novel hum
6	32	91.4	513	2 AAR99314	Aar99314 Rat MLPTP
7	32	91.4	613	7 ADL16248	Adl16248 Rat prote
8	32	91.4	682	6 ABU28623	Abu28623 Protein e
9	32	91.4	833	4 ABG09197	Abg09197 Novel hum
10	32	91.4	973	4 ABG15797	Abg15797 Novel hum
11	32	91.4	1400	3 ADC01365	Adc01365 Enterocae
12	32	91.4	1426	3 AAB15983	Aab15983 E. coli p
13	32	91.4	1426	6 ABU14693	Abu14693 Protein e
14	31	88.6	7	3 AAB02895	Aab02895 Nucleotid
15	31	88.6	7	7 ADF61985	Adf61985 Zinc fing
16	31	88.6	7	7 ADJ98427	Adj98427 Zinc fing
17	31	88.6	234	3 AAB51882	Aab51882 Gene 5 hu
18	31	88.6	251	4 AAE10130	Aae10130 Streptomy
19	31	88.6	263	4 AAB59376	Aab59376 Human pro
20	31	88.6	299	3 AAB34205	Aab34205 Gene 41 h
21	31	88.6	322	4 AAG78274	Aag78274 Human SIL
22	31	88.6	365	4 ADM19697	Adm19697 Protein e
23	31	88.6	393	5 AAB57822	Abp57822 Human RNA
24	31	88.6	393	6 ADA54298	Ada54298 Human pro
25	31	88.6	495	7 ABO74573	AbO74573 Pseudomon

26	31	88.6	518	4 AAB59241	Aab59241 SHP-1 pro
27	31	88.6	556	7 ADG74664	Adg74664 Human kin
28	31	88.6	556	8 ABM83634	Abm83634 Human dia
29	31	88.6	595	2 AAR99312	Aar99312 Human SH-
30	31	88.6	595	4 AAB59236	Aab59236 SHP-1 act
31	31	88.6	595	4 AAB59234	Aab59234 SHP-1 act
32	31	88.6	595	4 AAB59235	Aab59235 SHP-1 act
33	31	88.6	595	4 AAB59228	Aab59228 SHP-1 act
34	31	88.6	595	4 AAB59233	Aab59233 SHP-1 act
35	31	88.6	595	4 AAB59216	Aab59216 SHP-1 pro
36	31	88.6	595	4 AAB59239	Aab59239 SHP-1 E74
37	31	88.6	595	4 AAB59237	Aab59237 SHP-1 act
38	31	88.6	595	4 AAB59231	Aab59231 SHP-1 act
39	31	88.6	595	4 AAB59232	Aab59232 SHP-1 act
40	31	88.6	595	4 AAB59230	Aab59230 SHP-1 act
41	31	88.6	595	4 AAB59238	Aab59238 SHP-1 D59
42	31	88.6	595	4 AAB59229	Aab59229 SHP-1 act
43	31	88.6	595	7 ADL16238	Adl16238 Human pro
44	31	88.6	595	7 ADL16242	Adl16242 Human pro
45	31	88.6	595	8 ADH50796	Adh50796 Human SHP

ALIGNMENTS

RESULT 1	
AAB02894	
ID AAB02894 standard; peptide; 7 AA.	
XX	
AC AAB02894;	
XX	
DT 18-SEP-2000 (first entry)	
XX	
DE Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:35.	
XX	
KW Zinc finger domain; alpha helix; nucleotide binding; DNA binding;	
KW polyacyl protein; asymmetric target recognition;	
KW gene specific transcriptional regulator; gene activator; gene repressor;	
KW transcriptional switch; oncogene; erbA-2; cancer; tumour; gene therapy;	
KW transgenic animal; antiviral; anticancer; diagnosis.	
XX	
OS Synthetic.	
XX	
PN WO200023464-A2.	
XX	
PD 27-APR-2000.	
XX	
PF 14-OCT-1999; 99WO-EP007742.	
XX	
PR 16-OCT-1998; 98US-00173941.	
XX	
PA (NOVS) NOVARTIS AG.	
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.	
PA (SCRI) SCRIPPS RES INST.	
PI Barbas CF;	
XX	
DR MPI; 2000-339648/29.	
XX	
PT Novel isolated and purified zinc finger nucleotide-binding proteins with	
PT specificity for GNM triplet sequences, useful in gene therapy and for	
PT regulating gene function.	
XX	
PS Disclosure; Fig 1; 48pp; English.	
XX	
CC The invention relates to zinc finger nucleotide-binding proteins which	
CC comprise 2-12, preferably 2-6, operatively linked motifs selected from	
CC sequences AAB02860-802875. Sequences AAB02860-802875 represent the alpha	
CC helical regions of zinc finger domains which specifically bind to target	
CC nucleotide triplets of the sequence 5'-GNM-3'. Such regions may be linked	
CC by the peptide linker TGRKP (AAB02970). The Cys2-His2 zinc finger motif	
CC is the most frequently utilised nucleic acid binding motif in eukaryotes,	
CC and constitutes a beta-beta-alpha fold. Nucleic acid recognition is	

```

RA Amsterdam A., Nissen R.M., Sun Z., Swindell E.C., Farrington S.,
RA Hopkins N.,
RT "Identification of 315 genes essential for early zebrafish
RT development."
RL Proc. Natl. Acad. Sci. U.S.A. 101:12792-12797(2004).
RV [2]
RP SEQUENCE FROM N.A.
RC TISSUE=whole;
RX MEDLINE=22388257; PubMed=1247932; DOI=10.1073/pnas.242603899;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Stachenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,
RA Diegleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.L., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwen P.J., McKenna K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Heltón E., Kettelman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butcherfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalios D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marz M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RV [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole;
RX Director MGC Project;
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
RL EMBL: AY648824; AAT68142.1; -
DR EMBL: BC076020; AAH76020.1; -
DR GO: GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR InterPro: IPR004854; UFD1.
DR Pfam: PF03152; UFD1; 1.
SQ SEQUENCE 308 AA; 35000 MW; F5A890BFAA31ADBA CRC64;

Query Match 94.6%; Score 35; DB 2; Length 308;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDKR 7
DB 251 DPSDKR 257

RESULT 3
Q6ZIG6 PRELIMINARY; PRT; 763 AA.
AC Q6ZIG6;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Putative z1pA.
GN Name=O1115_B01.25; Synonyms=O11073_F05.14;
OS Oryza sativa (Japonica cultivar-group) .
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Euphorbiales; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrharioideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsunoto T., Yamamoto K.,
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP004000; BAD15444.1; -
DR EMBL: AP003990; BAD15414.1; -
SQ SEQUENCE 763 AA; 86146 MW; B5787A807577F693 CRC64;

Query Match 94.6%; Score 35; DB 2; Length 763;

```

Best Local Similarity	85.7%	Pred. No. 1.6e+02;			
Matches	6;	Conservative	1;	Mismatches	0;
				Indels	0;
				Gaps	0;
Oy	1	DPSDLR	7		
Db	653	DPSDLR	659		
RESULT 4					
O9RZAS					
ID	O9RZAS	PRELIMINARY;	PRT;	149 AA.	
AC	O9RZAS:				
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)			
DE	Response regulator.				
GN	OrderedLocusNames=DAA0049;				
OS	Deinococcus radiodurans.				
OC	Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;				
CC	Deinococcaceae; Deinococcus.				
OX	NCBI_TaxID=1299;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=RI / ATCC 13939 / DSM 20539 / NCIB 9279;				
RX	MEDLINE=20036896; PubMed=10567266; DOI=10.1126/science.286.5444.1571;				
RA	White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,				
RA	Doonan R.J., Haft D.H., Gwinn M.B., Nelson W.C., Richardson D.L.,				
RA	Moffatt K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,				
RA	Vamathevan J.J., Lam P., McDonald L.A., Ueberlack T.R., Zalewski C.,				
RA	Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,				
RA	Ketchum K.A., Nelson K.E., Salzberg S.L., Smith H.O., Venter J.C.,				
RA	Fraser C.M.;				
RT	"Genome sequence of the radioresistant bacterium Deinococcus				
RT	radiodurans R1."				
RL	Science 286:1571-1577(1999).				
DR	EMBL; AE001862; AAF12262.1; -.				
DR	PIR; C75598; C75598.				
DR	HSSP; Q5169; 113C.				
DR	TIGR; DRA0049; -.				
DR	GO; GO:0003677; F:DNA binding; IEA.				
DR	GO; GO:0000156; F:two-component response regulator activity; IEA.				
DR	GO; GO:0007600; F:sensory perception; IEA.				
DR	GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.				
DR	InterPro; IPR011006; Chey like.				
DR	InterPro; IPR001789; Response_reg.				
DR	Pfam; PF00072; Response_reg; 1.				
DR	ProDom; PD000039; Response_reg; 1.				
DR	SMART; SM00448; REC; 1.				
DR	PROSITE; PS00110; RESPONSE_REGULATORY; 1.				
DR	Complete proteome; Phosphorylation; Sensory transduction.				
KW	SEQUENCE 149 AA; 16604 MW; 6B30AFEB4957246A CRC64;				
SQL					
Query Match		91.9%;	Score 34;	DB 2;	Length 149;
Best Local Similarity	85.7%;	Pred. No. 42;			
Matches	6;	Conservative	1;	Mismatches	0;
				Indels	0;
				Gaps	0;
Oy	1	DPSDLR	7		
Db	100	DPSDLR	106		
RESULT 5					
O9RX00					
ID	O9RX00	PRELIMINARY;	PRT;	168 AA.	
AC	O9RX00:				
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)			
DE	Hypothetical protein DR0259.				
GN	OrderedLocusNames=DR0259;				
OS	Deinococcus radiodurans.				
OC	Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;				
CC	Deinococcaceae; Deinococcus.				

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:15:35 ; Search time 69.1667 seconds
(without alignments)

51.825 Million cell updates/sec

Title: US-10-646-919-34

Perfect score: 37

Sequence: 1 DPSDLKR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : uniprot_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	100.0	727	2	Q9CPJ7 pasteurella
2	35	94.6	308	2	Q6DRD5 brachydanio
3	35	94.6	763	2	Q6ZIG6 oryza sativ
4	34	91.9	149	2	Q9RZAS delinococcus
5	34	91.9	168	2	Q9RXQ0 delinococcus
6	34	91.9	277	2	Q9RUD3 arabidopsis
7	34	91.9	392	2	Q9YX73 aeropyrum p
8	34	91.9	798	2	Q75AG7 ashyba goes
9	34	91.9	812	2	Q6CL61 kluyveromyc
10	34	91.9	985	2	Q04387 saccharomyc
11	34	91.9	1032	2	Q80491 arabidopsis
12	34	91.9	1035	2	Q6N6N9 rhodospheudo
13	34	91.9	1116	2	Q8CHD2 mus musculu
14	34	91.9	1163	2	Q94872 homo sapien
15	34	91.9	1585	1	I4G3 HUMAN
16	34	91.9	1623	2	Q6RZX2 equus cabal
17	33	89.2	111	2	Q9YF60 aeropyrum p
18	33	89.2	153	2	Q9RTU3 delinococcus
19	33	89.2	210	2	Q6R2U7 musa acumina
20	33	89.2	215	2	Q22845 arabidopsis
21	33	89.2	265	2	Q8VC24 mus musculu
22	33	89.2	309	2	Q8MTY4 m cytochrome
23	33	89.2	335	2	Q22595 griffonia s
24	33	89.2	356	2	Q8LYD0 methanopyru
25	33	89.2	412	1	KAPR SCHPO
26	33	89.2	1353	2	Q62UM1 burkholderi
27	33	89.2	1356	2	Q63V28 burkholderi
28	32	86.5	138	2	Q8P926 leptospira
29	32	86.5	186	2	Q8KV86 unclutured
30	32	86.5	268	2	Q978S4 thermoplasma
31	32	86.5	304	2	Q9CU23 mus musculu

32	32	86.5	310	2	Q984W5 rhizobium 1
33	32	86.5	334	1	G3P PYRAB
34	32	86.5	334	1	G3P PYRFU
35	32	86.5	334	1	G3P PYRMO
36	32	86.5	336	1	SYW_GLOVI
37	32	86.5	340	2	Q9E299
38	32	86.5	344	2	Q9A9F8
39	32	86.5	354	1	SYW_AGR75
40	32	86.5	354	1	SYW_RHIME
41	32	86.5	355	1	SYW_BRUME
42	32	86.5	355	1	SYW_BRUSU
43	32	86.5	355	1	SYW_RHTLO
44	32	86.5	358	2	Q8C915
45	32	86.5	381	2	Q99XN6

ALIGNMENTS

```

RESULT 1
ID Q9CPJ7 PRELIMINARY; PRT; 727 AA.
AC Q9CPJ7;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Pthr.
GN Name=pfhr; OrderedLocusNames=PM0040;
OS Pasteurella multocida;
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
NCBI_TaxID=747;
RP [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Em70;
RX MEDLINE=21145866; PubMed=11248100; DOI=10.1073/pnas.051634598;
RA May B.U., Zhang Q., Li L.L., Paustian M.L., Whitlam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70." ;
RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006037; AK02124.1; -.
DR GO; GO:0019867; C:outer membrane; IEA.
DR GO; GO:0048772; F:receptor activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
KW Complete proteome.
SQ SEQUENCE 727 AA; 81332 MW; 714A089F8F60ADEB CRC64;

Query Match 100.0%; Score 37; DB 2; Length 727;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
DB 126 DPSDLKR 132

RESULT 2
ID Q6DRD5 PRELIMINARY; PRT; 308 AA.
AC Q6DRD5;
DT 25-OCT-2004 (TREMBLrel. 28, Created)
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Ubiquitin fusion degradation 1-like.
GN Name=utd1;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
NCBI_TaxID=7955;
RX NCB1
RX [1]
RX SEQUENCE FROM N.A.
RX PubMed=15256591; DOI=10.1073/pnas.0403929101;

```

A:Map position: 1

Query Match 91.9%; Score 34; DB 2; Length 168;
 Best Local Similarity 85.7%; Pred. No. 9.1;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
 |||||:
 DB 93 DPSDLKR 99

RESULT 3

D72511

probable hexosyltransferase (RC 2.4.1.-) APE2066 [similarity] - Aetopyrum pernix (strain N) alternate names: probable N-acetylglucosaminyl-1-phosphatidylinositol biosynthetic prote

C/Species: Aetopyrum pernix
 C/Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 09-Jul-2004

C/Accession: D72511
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Maeda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999
 A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aetopyr

A/Reference number: A72450; PMID:99310339; PMID:10382966

A/Accession: D72511

A/Molecule type: DNA

A/Residues: 1-339 <RAW>

A/Cross-references: UNIPROT:Q9YAF3; DDBJ:AP000063; NID:G5105654; PIDN:BAA81076.1; PID:G5

A/Experimental source: strain K1

C/Genetics:

C/Superfamily: probable hexosyltransferase ytxN

C/Keywords: glycosyltransferase; hexosyltransferase

Query Match 91.9%; Score 34; DB 1; Length 392;
 Best Local Similarity 85.7%; Pred. No. 22;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
 |||||:
 DB 91 DPSDLKR 97

RESULT 4

S15965

hypothetical protein 6 - yeast (Saccharomyces kluyveri) plasmid pSKL
 C/Species: Saccharomyces kluyveri
 C/Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004

C/Accession: S15965

R:Hishinuma, F.; Hirai, K.

Mol. Gen. Genet. 226, 97-106, 1991

A/Title: Genome organization of the linear plasmid, pSKL, isolated from Saccharomyces kl

A/Reference number: S15960; PMID:91238725; PMID:2034232

A/Accession: S15965

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-985 <HIS>

A/Cross-references: UNIPROT:Q04387; EMBL:X54850; NID:G4868; PIDN:CAA38625.1; PID:G4874

C/Genetics:

A/Genome: plasmid pSKL

Query Match 91.9%; Score 34; DB 2; Length 985;
 Best Local Similarity 85.7%; Pred. No. 59;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
 |||||:
 DB 705 DPSDLKR 711

RESULT 5

B86224

hypothetical protein (imported) - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C/Accession: B86224

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Comn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.,

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lures, J.S.; Maiti, R.; Marshall,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; PMID:21016719; PMID:11130712

A/Accession: B86224

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-1032 <STO>

A/Cross-references: UNIPROT:O80491; GB:AE005172; NID:G3249113; PIDN:AAC24096.1; GSPDB:GN

C/Genetics:

A/Map position: 1

Query Match 91.9%; Score 34; DB 2; Length 1032;
 Best Local Similarity 85.7%; Pred. No. 62;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
 |||||:
 DB 51 DPSDLKR 57

RESULT 6

D72730

hypothetical protein APE0381 - Aetopyrum pernix (strain K1)
 C/Species: Aetopyrum pernix
 C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004

C/Accession: D72730

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Maeda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K

DNA Res. 6, 83-101, 1999

A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aetopyr

A/Reference number: A72450; PMID:99310339; PMID:10382966

A/Accession: D72730

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-111 <RAW>

A/Cross-references: UNIPROT:Q9YF60; DDBJ:AP000059; NID:G5103911; PIDN:BAA79336.1; PID:dl

A/Experimental source: strain K1

C/Genetics:

A/Genome: APE0381

Query Match 89.2%; Score 33; DB 2; Length 111;
 Best Local Similarity 85.7%; Pred. No. 9.6;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
 |||||:
 DB 82 DPSDLKR 88

RESULT 7

B75367

hypothetical protein - Deinococcus radiodurans (strain R1)
 C/Species: Deinococcus radiodurans
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C/Accession: B75367

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A/Reference number: A75250; PMID:20036896; PMID:10567266

A/Accession: B75367

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:32:16 ; Search time 14.5 Seconds
(without alignments)
46.449 Million cell updates/sec

Title: US-10-646-919-34

Perfect score: 37

Sequence: 1 DPSDLKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	91.9	149	2 C75598	response regulator
2	34	91.9	168	2 C75541	conserved hypothe
3	34	91.9	392	1 D72511	probable hexosyltr
4	34	91.9	985	2 S15965	hypothetical prote
5	34	91.9	1032	2 B86224	hypothetical prote
6	33	89.2	111	2 D72730	hypothetical prote
7	33	89.2	153	2 B75367	hypothetical prote
8	33	89.2	169	2 D84864	probable calcium b
9	33	89.2	347	2 T39140	camp-dependent pro
10	33	89.2	411	2 S18634	cgsl protein - f18
11	32	86.5	334	1 DE0YG	glyceraldhyde-3-P
12	32	86.5	334	2 B87377	hypothetical prote
13	32	86.5	344	2 AE2619	tryptophan-tRNA 11
14	32	86.5	354	2 E97401	tryptophan-tRNA 11
15	32	86.5	383	2 AH3477	tryptophan-tRNA 11
16	32	86.5	385	2 S16321	light-induced prot
17	32	86.5	393	2 C87538	conserved hypothe
18	32	86.5	409	2 T41336	probable nitrogen
19	32	86.5	855	2 T43431	alpha-glucan synth
20	32	86.5	2352	2 F70514	hypothetical prote
21	31	83.8	145	2 T26085	hypothetical prote
22	31	83.8	312	2 T47621	bZIP transcription
23	31	83.8	403	2 B83507	probable ATP-bindi
24	31	83.8	588	2 JC6773	dynamin intermediat
25	31	83.8	694	2 G90696	hypothetical prote
26	31	83.8	706	2 JC6329	yeast secretory pr
27	31	83.8	708	2 C85547	probable cytoplasm
28	31	83.8	720	2 T31629	hypothetical prote
29	31	83.8	732	2 T31629	hypothetical prote

30	31	83.8	1036	2 E96682	hypothetical prote
31	31	83.8	1158	2 T25082	hypothetical prote
32	30	81.1	150	2 H86194	hypothetical prote
33	30	81.1	153	2 D84706	hypothetical prote
34	30	81.1	155	2 H69029	mutator Muet relat
35	30	81.1	192	2 T21210	hypothetical prote
36	30	81.1	216	1 ZZZRBL	nodulation protein
37	30	81.1	271	2 T31314	hypothetical prote
38	30	81.1	277	2 F71335	probable A/G-speci
39	30	81.1	281	2 S33023	probable uracil-DN
40	30	81.1	366	1 S35189	dihydrokaempferol
41	30	81.1	401	2 T41049	tyrosine-tRNA 11ga
42	30	81.1	440	2 G87444	8-amino-7-oxononan
43	30	81.1	451	2 T35914	probable carboxype
44	30	81.1	455	2 F71153	probable VI polysa
45	30	81.1	519	2 S66673	disulfide isomerases

ALIGNMENTS

RESULT 1
C75598
Response regulator - Deinococcus radiodurans (strain R1)
C/Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #ext_change 16-Aug-2004
C/Accession: C75598
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.; M
Science 286, 1571-1577, 1999
A/Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.
A/Reference number: A75250; PMID:20036896; PMID:10567266
A/Accession: C75598
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-149 <WHI>
A/Cross-references: UNIPROT:Q9R2A5; GB:AE001862; GB:AE001825; NID:96460468; PIDN:AAF122
A/Experimental source: strain R1
C/Genetics:
A/Gene: DR0049
A/Map position: 2
C/Superfamily: Signal transduction receiver (phosphoacceptor) protein, CheY type; respo

Query Match 91.9%; Score 34; DB 2; Length 149;
Best Local Similarity 85.7%; Pred. No. 8;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
Db 100 DPSDLKR 106
|||||
|||

RESULT 2
C75541
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C/Species: Deinococcus radiodurans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #ext_change 09-Jul-2004
C/Accession: C75541
R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.; M
Science 286, 1571-1577, 1999
A/Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.
A/Reference number: A75250; PMID:20036896; PMID:10567266
A/Accession: C75541
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-168 <WHI>
A/Cross-references: UNIPROT:Q9R0X0; GB:AE001867; GB:AE000513; NID:96457928; PIDN:AAF098
A/Experimental source: strain R1
C/Genetics:
A/Gene: DR0259

RESULT 2
US-10-437-963-173601
; Sequence 173601, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 173601
; LENGTH: 763
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_71622C.1.pep
US-10-437-963-173601

Query Match 94.6%; Score 35; DB 16; Length 763;
Best Local Similarity 85.7%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLR 7
|||:|
Db 653 DPSDLR 659

RESULT 3
US-10-425-115-359614
; Sequence 359614, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 359614
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_91140C.1.pep
US-10-425-115-359614

Query Match 91.9%; Score 34; DB 16; Length 112;
Best Local Similarity 85.7%; Pred. No. 57;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLR 7
|||:|
Db 48 DPTDLR 54

RESULT 4
US-10-424-599-220780
; Sequence 220780, Application US/10424599
; Publication No. US20040031072A1

; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 220780
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_41395C.1.pep
US-10-424-599-220780

Query Match 91.9%; Score 34; DB 15; Length 283;
Best Local Similarity 85.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLR 7
|||:|
Db 118 DPTDLR 124

RESULT 5
US-10-369-493-22940
; Sequence 22940, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22940
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Aeropyrum pernix
US-10-369-493-22940

Query Match 91.9%; Score 34; DB 15; Length 392;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLR 7
|||:|
Db 91 DPSDLR 97

RESULT 6
US-10-120-801-71
; Sequence 71, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shinkets, Richard
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehraban, Fuad

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:52:31 ; Search time 71.333 Seconds
(without alignments)
38.093 Million cell updates/sec

Title: US-10-646-919-34

Perfect score: 37
Sequence: 1 DPSDLKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1736639 seqs, 388188149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/prodata/2/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/2/pubppa/PCT_NEW_PUB.pep:*
3: /cgn2_6/prodata/2/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/prodata/2/pubppa/US06_PUBCOMB.pep:*
5: /cgn2_6/prodata/2/pubppa/US07_NEW_PUB.pep:*
6: /cgn2_6/prodata/2/pubppa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/prodata/2/pubppa/US08_NEW_PUB.pep:*
8: /cgn2_6/prodata/2/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/prodata/2/pubppa/US09_PUBCOMB.pep:*
10: /cgn2_6/prodata/2/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/prodata/2/pubppa/US09_PUBCOMB.pep:*
12: /cgn2_6/prodata/2/pubppa/US09_NEW_PUB.pep:*
13: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
14: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
15: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
16: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
17: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
18: /cgn2_6/prodata/2/pubppa/US10_PUBCOMB.pep:*
19: /cgn2_6/prodata/2/pubppa/US11_PUBCOMB.pep:*
20: /cgn2_6/prodata/2/pubppa/US11_NEW_PUB.pep:*
21: /cgn2_6/prodata/2/pubppa/US60_NEW_PUB.pep:*
22: /cgn2_6/prodata/2/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100.0	7	18	US-10-646-919-34	Sequence 34, Appl
2	94.6	763	16	US-10-437-963-173601	Sequence 173601, A
3	91.9	112	16	US-10-425-115-359614	Sequence 359614, A
4	91.9	283	15	US-10-424-599-220780	Sequence 220780, A
5	91.9	392	15	US-10-369-493-22940	Sequence 22940, A
6	91.9	901	15	US-10-120-801-71	Sequence 71, Appl
7	89.2	1163	15	US-10-120-801-70	Sequence 70, Appl
8	89.2	412	14	US-10-092-947A-41	Sequence 41, Appl
9	89.2	752	15	US-10-282-122A-49178	Sequence 49178, A
10	89.2	1319	15	US-10-282-122A-50503	Sequence 50503, A
11	86.5	131	16	US-10-425-115-190872	Sequence 190872, A

	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	
	32	32	32	32	32	32	32	32	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	31	
	86.5	86.5	86.5	86.5	86.5	86.5	86.5	86.5	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	83.8	
	137	137	137	137	137	137	137	137	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	106	
	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
	US-10-425-114-47736	US-10-425-115-338739	US-10-282-122A-74282	US-09-971-450-8	US-10-193-452-81	US-10-369-493-19467	US-10-437-963-120511	US-10-424-599-146899	US-10-424-599-204330	US-10-767-701-63052	US-10-424-599-168338	US-10-767-701-32650	US-10-425-114-42428	US-10-425-115-224817	US-10-424-599-220344	US-10-425-114-40223	US-10-424-599-156669	US-10-425-114-39771	US-10-437-963-161723	US-10-408-765A-968	US-10-408-765A-968	US-10-156-761-7674	US-10-156-761-7674	US-10-425-115-303434	US-10-425-114-52815	US-10-437-963-186038	US-10-425-114-64694	US-10-302-267-52	US-10-412-699B-812	US-10-424-599-259544	US-10-424-599-151645	US-10-437-963-157591	US-10-282-122A-65028	US-10-424-599-198122	US-10-425-115-348539
	Sequence 47736, A	Sequence 338739, A	Sequence 74282, A	Sequence 8, Appl	Sequence 81, Appl	Sequence 19467, A	Sequence 120511, A	Sequence 146899, A	Sequence 204330, A	Sequence 63052, A	Sequence 156669, A	Sequence 32650, A	Sequence 42428, A	Sequence 224817, A	Sequence 220344, A	Sequence 40223, A	Sequence 156669, A	Sequence 39771, A	Sequence 161723, A	Sequence 161726, A	Sequence 968, App	Sequence 7674, App	Sequence 303434, A	Sequence 52815, A	Sequence 186038, A	Sequence 64694, A	Sequence 52, Appl	Sequence 812, App	Sequence 259544, A	Sequence 161645, A	Sequence 127591, A	Sequence 65028, A	Sequence 198122, A	Sequence 348539, A	

ALIGNMENTS

RESULT 1
US-10-646-919-34
Sequence 34, Application US/10646919
Publication No. US20050148075A1
GENERAL INFORMATION:
APPLICANT: Barbae, Carlos F.
TITLE OF INVENTION: Zinc-finger binding domains for GNN
FILE REFERENCE: TSRI 645.2C1
CURRENT APPLICATION NUMBER: US/10/646,919
CURRENT FILING DATE: 2003-08-21
PRIOR APPLICATION NUMBER: US 09/494,190
PRIOR FILING DATE: 2000-01-28
PRIOR APPLICATION NUMBER: PCT/EP99/07742
PRIOR FILING DATE: 1999-10-14
PRIOR APPLICATION NUMBER: US 09/173,941
PRIOR FILING DATE: 1998-10-16
NUMBER OF SEQ ID NOS: 129
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 34
LENGTH: 7
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthesized
US-10-646-919-34

Query Match 100.0%; Score 37; DB 18; Length 7;
Best local similarity 100.0%; Pred. No. 1.6e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
DB 1 DPSDLKR 7

FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:phage display
; OTHER INFORMATION: Selected and mutagenized
US-09-494-190-34

Query Match 100.0%; Score 37; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 4,1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
|||
Db 1 DPSDLKR 7

RESULT 3
US-09-543-681A-6003
; Sequence 6003, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:

; APPLICANT: GARY BRETTON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6003
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6003

Query Match 91.9%; Score 34; DB 4; Length 364;
Best Local Similarity 85.7%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
|||
Db 215 DPSDLKR 221

RESULT 4
US-09-248-796A-26549
; Sequence 26549, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196,132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 26549
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-26549

Query Match 86.5%; Score 32; DB 4; Length 311;
Best Local Similarity 85.7%; Pred. No. 88;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
|||
Db 235 DPSDLKR 241

RESULT 5
US-09-949-016-6838
; Sequence 6838, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6838
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-6838

Query Match 83.8%; Score 31; DB 4; Length 396;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PSDLKR 7
|||
Db 315 PSDLKR 320

RESULT 6
US-09-949-016-10646
; Sequence 10646, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10646
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10646

Query Match 83.8%; Score 31; DB 4; Length 409;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PSDLKR 7
|||
Db 328 PSDLKR 333

RESULT 7
US-09-252-991A-22913
; Sequence 22913, Application US/09252991A
; Patent No. 6551795

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:17:05 ; Search time 21.6667 Seconds
(without alignments)
24.117 Million cell updates/sec

Title: US-10-646-919-34

Perfect score: 37

Sequence: 1 DSDDKR 7

Scoring table: BLOSUM62

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/iaa/5A_COMB.pep:*

2: /cgn2_6/prodata/1/iaa/5B_COMB.pep:*

3: /cgn2_6/prodata/1/iaa/6A_COMB.pep:*

4: /cgn2_6/prodata/1/iaa/6B_COMB.pep:*

5: /cgn2_6/prodata/1/iaa/PCTUS_COMB.pep:*

6: /cgn2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	100.0	7	US-09-173-941-34	Sequence 34, Appl
2	37	100.0	7	US-09-494-190-34	Sequence 34, Appl
3	34	91.9	364	US-09-543-681A-6003	Sequence 6003, Ap
4	32	86.5	311	US-09-248-796A-26549	Sequence 26549, A
5	31	83.8	396	US-09-949-016-6838	Sequence 6838, Ap
6	31	83.8	409	US-09-949-016-10646	Sequence 10646, A
7	31	83.8	440	US-09-252-991A-22913	Sequence 22913, A
8	31	83.8	475	US-09-252-991A-31373	Sequence 31373, A
9	31	83.8	599	US-09-252-991A-19543	Sequence 19543, A
10	31	83.8	708	US-09-643-597-369	Sequence 369, Ap
11	31	83.8	723	US-09-328-352-6928	Sequence 19543, A
12	31	83.8	818	US-09-252-991A-16691	Sequence 16691, A
13	31	83.8	906	US-09-252-991A-32715	Sequence 32715, A
14	31	83.8	1089	US-09-252-991A-20334	Sequence 20334, A
15	31	83.8	7	US-09-173-941-33	Sequence 33, Appl
16	30	81.1	7	US-09-494-190-33	Sequence 33, Appl
17	30	81.1	108	US-09-540-236-3176	Sequence 3176, Ap
18	30	81.1	271	US-09-248-796A-16882	Sequence 16882, A
19	30	81.1	271	US-09-408-020-16	Sequence 16, Appl
20	30	81.1	272	US-09-177-165A-27	Sequence 27, Appl
21	30	81.1	315	US-09-252-991A-19219	Sequence 19219, A
22	30	81.1	366	US-09-638-715-2	Sequence 2, Appl1
23	30	81.1	366	US-09-638-715-4	Sequence 2, Appl1
24	30	81.1	366	US-10-060-509-2	Sequence 2, Appl1
25	30	81.1	366	US-10-060-509-4	Sequence 4, Appl1
26	30	81.1	366	US-10-060-506-2	Sequence 2, Appl1
27	30	81.1	366	US-10-060-506-4	Sequence 4, Appl1

28	30	81.1	497	4	US-09-252-991A-30012	Sequence 30012, A
29	30	81.1	788	4	US-09-252-991A-17380	Sequence 17380, A
30	30	81.1	815	4	US-09-177-165A-24	Sequence 24, Appl
31	30	81.1	912	4	US-09-252-991A-24163	Sequence 24163, A
32	29	78.4	90	4	US-09-270-767-32706	Sequence 32706, A
33	29	78.4	118	4	US-09-248-796A-28195	Sequence 28195, A
34	29	78.4	185	4	US-09-562-737-115	Sequence 115, App
35	29	78.4	209	2	US-08-870-518-10	Sequence 9, Appl
36	29	78.4	328	4	US-09-679-279-9	Sequence 6618, Ap
37	29	78.4	370	4	US-09-134-000C-6618	Sequence 23259, A
38	29	78.4	378	4	US-09-252-991A-23259	Sequence 2, Appl1
39	29	78.4	459	1	US-08-630-592-2	Sequence 2, Appl1
40	29	78.4	459	1	US-08-714-991-2	Sequence 2, Appl1
41	29	78.4	459	2	US-08-870-518-1	Sequence 2, Appl1
42	29	78.4	459	2	US-09-032-365A-2	Sequence 1, Appl1
43	29	78.4	460	1	US-08-630-592-7	Sequence 7, Appl1
44	29	78.4	460	1	US-08-714-991-7	Sequence 7, Appl1
45	29	78.4	460	3	US-09-032-365A-8	Sequence 8, Appl1

ALIGNMENTS

```
RESULT 1
US-09-173-941-34
; Sequence 34, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV00815
; CURRENT APPLICATION NUMBER: US/09/173,941
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
; OTHER INFORMATION: codon binding sequence
US-09-173-941-34

Query Match      100.0%; Score 37; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DSDDKR 7
      |||||
Db      1 DSDDKR 7

RESULT 2
US-09-494-190-34
; Sequence 34, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 34
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
OS Bacteria.
PN US2003233675-A1.
PD 18-DEC-2003.
PF 20-FEB-2003; 2003US-00369493.
PR 21-FEB-2002; 2002US-0360039P.
XX (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.
XX (SLAT/) SLATER S C.
XX (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX New recombinant DNA construct comprising a promoter positioned to provide
XX PT for expression of a polynucleotide encoding a polypeptide from a
XX PT microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 22940; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
XX CC promoter functional in a plant cell, where the promoter is positioned to
XX CC provide for expression of a polynucleotide encoding a polypeptide from a
XX CC microbial source. The invention also relates to a transformed plant
XX CC comprising the recombinant DNA construct and a method of producing a
XX CC transformed plant having an improved property. The plant is a crop plant
XX CC such as maize or soybean. The method of producing a transformed plant
XX CC having an improved property comprises transforming a plant with the
XX CC recombinant DNA construct and growing the transformed plant, where the
XX CC polynucleotide or polypeptide is useful for improving plant properties.
XX CC The recombinant DNA construct is useful for producing plants with
XX CC improved plant properties, e.g. improved cold, heat or drought tolerance,
XX CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX CC increased resistance to plant disease, better growth rate by modification
XX CC of the cell cycle pathway with plant growth regulators, increased rate of
XX CC homologous recombination, modified seed oil or protein yield and/or
XX CC content, improved yield by modification of carbohydrate, nitrogen or
XX CC phosphorus use and/or uptake, by modification of photosynthesis or by
XX CC providing improved plant growth and development under at least one stress
XX CC condition, improved lignin production or improved galactomannan
XX CC production. This sequence represents a bacterial polypeptide used in the
XX CC scope of the invention. Note: The sequence data for this patent did not
XX CC form part of the printed specification but was obtained in electronic
XX CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX SQ Sequence 392 AA;
XX
XX Query Match 91.9%; Score 34; DB 8; Length 392;
XX Best Local Similarity 85.7%; Pred. No. 1.6e+02;
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 DPSDLKR 7
XX |||||:
XX Db 91 DPSDLRR 97
XX
XX RESULT 7
XX ADG39839 standard; protein; 901 AA.
XX
XX AC ADG39839;

XX
XX 26-FEB-2004 (first entry)
XX DT
XX DE Protein similar to human NOV8 #5.
XX
XX Human; NOVX; cardiomyopathy; atherosclerosis; hypertension;
KW congenital heart defect; pulmonary stenosis; scleroderma; obesity;
KW metabolic disturbance; obesity; transplantation; adrenoleukodystrophy;
KW congenital adrenal hyperplasia; prostate cancer; diabetes;
KW metabolic disorder; neoplasm; adenocarcinoma; fertility; haemophilia;
KW graft versus host disease; AIDS; bronchial asthma; Crohn's disease;
KW multiple sclerosis; infectious disease; anorexia;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW immune disorder; hematopoietic disorder; dyslipidaemia;
KW wasting disorder.
XX
XX Homo sapiens.
XX OS
XX US2003203843-A1.
XX PN
XX 30-OCT-2003.
XX PD
XX
XX 11-APR-2002; 2002US-00120801.
XX PF
XX 20-APR-2001; 2001US-0285609P.
XX PR 23-APR-2001; 2001US-0285748B.
XX PR 24-APR-2001; 2001US-0286068P.
XX PR 25-APR-2001; 2001US-0286292P.
XX PR 03-MAY-2001; 2001US-0288334P.
XX PR 16-MAY-2001; 2001US-0291241P.
XX PR 14-SEP-2001; 2001US-0322284P.
XX
XX (PENA/) PENA C E A.
XX (GUOX/) GUO X.
XX (SHIM/) SHIMKETS R A.
XX (PADI/) PADIGARU M.
XX (KEKU/) KERUDA R.
XX (SPYT/) SPYTEK K A.
XX (MEHR/) MEHRABAN F.
XX (TOPP/) TOPPER J N.
XX (MALY/) MALYANKAR U M.
XX (WASS/) WASSERMAN S M.
XX (EDIN/) EDINGER S R.
XX (SMIT/) SMITHSON G.
XX (GUNT/) GUNTHER E.
XX (KOMU/) KOMUVES L.
XX
XX Pena CE, Guo X, Shimkets RA, Padigaru M, Kekuda R, Spytsek KA;
XX Mehraban F, Topper JN, Malyankar UM, Wasserman SM, Edinger SR;
XX Smithson G, Gunther E, Komuves L;
XX WPI; 2003-900671/82.
XX
XX New NOVX polypeptides and nucleic acids, useful for diagnosing or
XX PT treating e.g. cardiomyopathy, atherosclerosis, hypertension, scleroderma,
XX PT obesity, prostate cancer, AIDS, bronchial asthma, Crohn's disease, or
XX PT multiple sclerosis.
XX
XX Disclosure; SEQ ID NO 71; 215pp; English.
XX
XX The invention relates to a new isolated polypeptide comprising an amino
XX CC acid sequence selected from 17 fully defined human NOVX sequences (even
XX CC SEQ ID NOS between ADG39770 and ADG39802), a mature form of the NOVX
XX CC amino acid or a variant of NOVX, where one or more amino acid residue in
XX CC the variant differs in no more than 15% of the amino acid residues of
XX CC NOVX. Also included are an isolated nucleic acid (NA) molecule
XX CC comprising a nucleic acid sequence encoding a NOVX polypeptide above
XX CC (odd SEQ ID NOS between ADG39769 and ADG39801), a nucleic acid fragment
XX CC encoding at least a portion of a NOVX polypeptide and a complement of
XX CC NOVX NA, a vector comprising NOVX NA, a cell comprising the vector, an
XX CC antibody that immunospecifically binds to NOVX, a method for determining
XX CC the presence or amount of NOVX or NOVX NA in a sample, a method of
XX CC identifying an agent that binds to NOVX, a method for identifying an

RESULT 4

ABG20437 standard; protein; 178 AA.

XX ABG20437;

XX 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #20428.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX MO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Dmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS84624.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

PT Claim 20; SEQ ID NO 50796; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences)

XX Sequence 178 AA;

XX Query Match 91.9%; Score 34; DB 4; Length 178;

XX Best Local Similarity 85.7%; Pred. No. 65;

XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 DPSDLKR 7

DB 100 DPSDLKK 106

RESULT 5

ADF05718 standard; protein; 364 AA.

XX ADF05718;

XX 12-FEB-2004 (first entry)

DE Bacterial polypeptide #1831.

KW Proteus mirabilis infection; bacterial infection; antibacterial;
KW immunostimulant.

XX Proteus mirabilis.

XX US6605709-B1.

XX 12-AUG-2003.

XX 05-APR-2000; 2000US-00543681.

XX 09-APR-1999; 99US-0128706P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Breton GL;

XX WPI; 2003-895291/82.

XX N-PSDB; ADF01546.

PT New Proteus mirabilis polypeptides and polynucleotides, useful as
PT reagents for diagnosis of bacterial disease, as components of
PT antibacterial vaccines, as targets for antibacterial drugs, or as
PT biocontrol agents for plants.

PT Disclosure, SEQ ID NO 6003; 870bp; English.

The invention relates to new Proteus mirabilis polypeptides and polynucleotides. The invention also relates to antibodies against the polypeptides, methods for producing the polypeptides, a method of generating vaccines for immunising an individual against P. mirabilis, a method for evaluating a compound for the ability to bind a P. mirabilis polypeptide and a method for screening test compounds for anti-bacterial activity. The polypeptides and polynucleotides are useful as molecular targets for diagnosing, preventing and treating pathological conditions resulting from bacterial infection, as reagents for diagnosis of bacterial diseases, as components of antibacterial vaccines, as targets for antibacterial drugs or as bio-control agents for plants. This sequence represents a Proteus mirabilis polypeptide of the invention.

XX Sequence 364 AA;

XX Query Match 91.9%; Score 34; DB 7; Length 364;

XX Best Local Similarity 85.7%; Pred. No. 1,5e+02;

XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 DPSDLKR 7

DB 215 DPNDLKR 221

RESULT 6

ADS44510 standard; protein; 392 AA.

XX ADS44510;

XX 02-DEC-2004 (first entry)

DE Bacterial polypeptide #22940.

KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;

CC achieved through specific contacts from side chains of amino acid
CC residues in the alpha helix. Each zinc finger can recognise a subunit of
CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains
CC allows the recognition of extended contiguous asymmetric DNA sequences.
CC For example, a synthetic polynucleotide protein containing six zinc finger
CC domains can recognise an 18 bp sequence, and such proteins are
CC potentially highly gene-specific. The novel nucleotide-binding zinc
CC finger proteins may therefore be used in the development of artificial
CC gene-specific transcriptional regulators. Such transcriptional switches
CC may be used to regulate the expression of oncogenes such as erbB-2,
CC overexpression of which is involved in malignant transformation. The
CC proteins are therefore useful in the treatment of cancers, and may also
CC be used to activate genes involved in fighting diseases, and to treat
CC viral infections by inhibiting the synthesis of viral gene products. They
CC may be used in DNA-based diagnostic applications. The proteins may also
CC be used in producing functional gene knockout or activation in
CC heterozygous transgenic animals. Proteins of the invention can
CC discriminate between sequences which have a single base difference. This
CC is manifested in a >100-fold decrease in affinity for the variant
CC sequence. Gene activation and repression can be achieved by targeting
CC within the gene transcript, suggesting that information obtained from
CC expressed sequence tags may be sufficient for the construction of gene
CC switches. Sequences AAB02876-B02869 represent zinc finger alpha helix
CC phage library peptides disclosed in the invention
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 37; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
Db 1 DPSDLKR 7

RESULT 2
ADP61983
ID ADF61983 standard; peptide; 7 AA.

AC ADF61983;

DT 12-FEB-2004 (first entry)

XX zinc finger binding region #34.

XX zinc finger-nucleotide binding polypeptide; expression regulation;

XX zinc finger binding region.

OS Synthetic.

PN US6610512-B1.

PD 26-AUG-2003.

PF 28-JAN-2000; 2000US-00494190.

PR 16-OCT-1998; 98US-00173941.

PR 14-OCT-1999; 99WO-EP007742.

PA (SCRI) SCRIPPS RES INST.

PI Barbas CF;

XX MPI; 2003-800134/75.

Regulating expression of nucleotide sequence that contains sequence 5'-
(GNN)n-3', comprising exposing nucleotide sequence to composition
comprising isolated and purified zinc finger-nucleotide binding
polypeptide.
Disclosure; SEQ ID NO 34; 46pp; English.

CC The invention relates to an expression of nucleotide sequence that
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by
CC exposing nucleotide sequence to composition comprising isolated and
CC purified zinc finger-nucleotide binding polypeptide that contains a
CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Ieu-Val-
CC Arg. The invention is used in the regulation of the expression of the
CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present
CC sequence represents the amino acid sequence of a zinc finger binding
CC region.
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 37; DB 7; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
Db 1 DPSDLKR 7

RESULT 3
ADJ98425
ID ADJ98425 standard; peptide; 7 AA.

AC ADJ98425;

DT 06-MAY-2004 (first entry)

DE zinc finger DNA binding peptide #96.

XX library; multimeric DNA binding polypeptide;

XX zinc finger DNA binding peptide; gene expression silencing;

XX gene expression enhancement.

XX Unidentified.

OS WO200306828-A2.

PN 14-AUG-2003.

PD 07-FEB-2003; 2003WO-US003705.

PR 07-FEB-2002; 2002US-0354981P.

XX (SCRI) SCRIPPS RES INST.

XX Barbas CF, Blancafort P;

PI MPI; 2003-731499/69.

New zinc finger library of multimeric DNA binding polypeptides, useful
for sterically occluding the binding site of a natural transcription
factor, and enhancing or silencing target gene expression.

PS Disclosure; Fig 20; 64pp; English.

XX The invention comprises a library of multimeric DNA binding polypeptides
XX (e.g. zinc finger DNA binding polypeptides). The zinc finger binding
XX polypeptides of the invention are useful for sterically occluding the
XX binding site of a natural transcription factor and enhancing or silencing
XX target gene expression. The present amino acid sequence represents a zinc
XX finger DNA binding peptide of the invention.
XX

SQ Sequence 7 AA;

Query Match 100.0%; Score 37; DB 7; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.8e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSDLKR 7
Db 1 DPSDLKR 7

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:14:30 ; Search time 82.5 Seconds
(without alignments)
32.816 Million cell updates/sec

Title: US-10-646-919-34

Perfect score: 37

Sequence: 1 DPSDKR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_16Dec04:*

1: geneseqp1980s:*

2: geneseqp1990s:*

3: geneseqp2000s:*

4: geneseqp2001s:*

5: geneseqp2002s:*

6: geneseqp2003as:*

7: geneseqp2003bs:*

8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the change being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	37	100.0	7	3	AAB02893	Aab02893 Nucleotid
2	37	100.0	7	7	ADP61983	Adf61983 Zinc fing
3	37	100.0	7	7	ADJ98425	Adj98425 Zinc fing
4	34	91.9	178	4	ABG20437	Abg20437 Novel hum
5	34	91.9	364	7	ADP05718	Adp05718 Bacterial
6	34	91.9	352	8	ADS44510	Ads44510 Bacterial
7	34	91.9	901	7	ADG39839	Adg39839 Protein s
8	34	91.9	1139	8	AD116228	Ad116228 Human nuc
9	34	91.9	1163	7	ADG39838	Adg39838 Protein s
10	34	91.9	1556	8	ADG97964	Adg97964 Human can
11	34	91.9	1584	3	AA184430	AA184430 Amino aci
12	34	91.9	1585	8	ADG97962	Adg97962 Human can
13	34	91.9	1756	4	ABG20440	Abg20440 Novel hum
14	33	89.2	169	3	AAC43188	Aac43188 Arabidops
15	33	89.2	169	3	AAG07957	Aag07957 Arabidops
16	33	89.2	215	3	AAG43187	Aag43187 Arabidops
17	33	89.2	215	3	AAG07956	Aag07956 Arabidops
18	33	89.2	215	8	ADN73189	Adn73189 Thale cte
19	33	89.2	309	7	ADC49305	Adc49305 Mouse inh
20	33	89.2	356	7	ADM25766	Adm25766 Hyperther
21	33	89.2	752	6	ABU21254	Abu21254 Protein e
22	33	89.2	1319	6	ABU22579	Abu22579 Protein e
23	32	86.5	179	6	ABU46358	Abu46358 Protein e
24	32	86.5	314	4	AAB96123	Aab96123 Putative
25	32	86.5	387	5	ABP25701	Abp25701 Streptoco

26	32	86.5	459	8	ADG44904	Adg44904 Mouse bac
27	32	86.5	468	5	AAE24155	Aae24155 Human 466
28	32	86.5	468	8	ADG44908	Adg44908 Human sac
29	32	86.5	571	8	ADP04715	Adp04715 Sea equit
30	32	86.5	676	8	ADG30434	Adg30434 Bacterial
31	31	83.8	158	4	ABG13584	Abg13584 Novel hum
32	31	83.8	205	4	ABG25323	Abg25323 Novel hum
33	31	83.8	298	7	ADM25590	Adm25590 Hyperther
34	31	83.8	357	7	ADJ69162	Adj69162 Human hea
35	31	83.8	396	6	ABH82854	Abh82854 SPO11 rel
36	31	83.8	396	6	ABH82855	Abh82855 SPO11 rel
37	31	83.8	403	8	AD161353	Ad161353 A. thalia
38	31	83.8	403	8	AD002359	Ado02359 Thalecres
39	31	83.8	417	4	AA063550	Aa063550 Propionib
40	31	83.8	417	6	ABM60069	Abm60069 Propionib
41	31	83.8	440	7	ABO74167	Ab074167 Pseudomon
42	31	83.8	448	6	ABP72283	Abp72283 N. gonorr
43	31	83.8	448	6	ABU37104	Abu37104 Protein e
44	31	83.8	475	7	ABO82627	Ab082627 Pseudomon
45	31	83.8	588	7	ADG20412	Adg20412 Pseudomon

ALIGNMENTS

RESULT 1
AAB02893
ID AAB02893 standard; peptide; 7 AA.
XX
AC AAB02893;
XX
DT 18-SEP-2000 (first entry)
XX
DE Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:34.
XX
KW Zinc finger domain; alpha helix; nucleotide binding; DNA binding;
KW polylactyl protein; asymmetric target recognition;
KW gene specific transcriptional regulator; gene activator; gene repressor;
KW transcriptional switch; oncogene; erbB-2; cancer; tumour; gene therapy;
KW transgenic animal; antiviral; anticancer; diagnosis.
XX
OS Synthetic.
XX
PN WO200023464-A2.
XX
PD 27-APR-2000.
XX
PF 14-OCT-1999; 99WO-EP007742.
XX
PR 16-OCT-1998; 98US-00173941.
XX
PA (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PA (SCRI) SCRIPPS RES INST.
PI Barbas CF;
XX
DR WPI, 2000-339648/29.
XX
PT Novel isolated and purified zinc finger nucleotide-binding proteins with
XX specificity for GNN triplet sequences, useful in gene therapy and for
XX regulating gene function.
XX
PS Disclosure, Fig 1; 48pp; English.
XX
CC The invention relates to zinc finger nucleotide-binding proteins which
XX comprise 2-12, preferably 2-6, operatively linked motifs selected from
XX sequences AAB02860-802875. Sequences AAB02860-802875 represent the alpha
XX helical regions of zinc finger domains which specifically bind to target
XX nucleotide triplets of the sequence 5'-GNN-3'. Such regions may be linked
XX by the peptide linker TGRKP (AAB02970). The Cys2-His2 zinc finger motif
XX is the most frequently utilised nucleic acid binding motif in eukaryotes,
XX and constitutes a beta-beta-alpha fold. Nucleic acid recognition is

Query Match 94.4%; Score 34; DB 2; Length 463;
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DPSTLKR 7
 |||||
 Db 52 DPSTIKR 58

RESULT 2

094KIT0 PRELIMINARY; PRT; 527 AA.

AC 094KIT0; PRELIMINARY; PRT; 527 AA.
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, last annotation update)
 DE Mitochondrial processing peptidase.
 OS Avicennia marina (Grey mangrove).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Lamiaceae; Acanthaceae; Acanthaceae incertae sedis; Avicennia.
 NCBI_TaxID=82927;

RA Parani M., Parida A.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; A363285; AAK51086.1; -.
 DR HSSP; P10507; 1HR6.
 DR MEROPS; M16.003; -.
 DR GO; GO:0004222; F:metalloendopeptidase activity; IEA.
 DR GO; GO:0005508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR001431; Insulinase-like.
 DR InterPro; IPR007863; Peptidase_M16_C.1.
 DR Pfam; PF00675; Peptidase_M16_C.1.
 DR Pfam; PF05193; Peptidase_M16_C.1.
 DR PROSITE; PS00143; INSULINASE; 1.
 SQ SEQUENCE 527 AA; 59239 MW; 5DA02B8BD04F9AF CRC64;

Query Match 94.4%; Score 34; DB 2; Length 527;
 Best Local Similarity 85.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DPSTLKR 7
 |||||
 Db 484 DPSTIKR 490

RESULT 3

08P0G5 PRELIMINARY; PRT; 792 AA.

AC 08P0G5; PRELIMINARY; PRT; 792 AA.
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, last annotation update)
 DE Putative internalin A.
 GN OrderedLocustNames=spym18_1373;
 OS Streptococcus pyogenes (serotype M18).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=186103;

RA [1] SEQUENCE FROM N.A.
 RC STRAIN=MGAS8232;
 RX MEDLINE=21927593; PubMed=11917108; DOI=10.1073/pnas.062526099;
 RA Smoot J.C., Barbhan K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
 RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
 RA Kapur V., Daly J.A., Veasy L.G., Nusser J.M.;
 RT "Genome sequence and comparative microarray analysis of serotype M18
 RT group A Streptococcus strains associated with acute rheumatic fever
 RT outbreaks";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).

DR EMBL; AE010057; AAL97968.1; -.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR007092; LRR_SD522.
 DR InterPro; IPR006270; Strept_his_triad.
 DR Pfam; PF00560; LRR_1_5.
 DR Pfam; PF04270; Strept_his_triad; 3.
 DR PRINTS; PR00019; LEURICHRPT.
 KW Complete proteome.
 SQ SEQUENCE 792 AA; 87440 MW; 5C4DDEBELBA991A7 CRC64;

Query Match 94.4%; Score 34; DB 2; Length 792;
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DPSTLKR 7
 |||||
 Db 299 DPSTIKR 305

RESULT 4

099276 PRELIMINARY; PRT; 792 AA.

AC 099276; PRELIMINARY; PRT; 792 AA.
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, last annotation update)
 DE Putative internalin A.
 GN Name=InLA; OrderedLocustNames=SPY1361;
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 NCBI_TaxID=1314;
 RN [1]

RA SEQUENCE FROM N.A.
 RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
 RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;
 RA Ferrer J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
 RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
 RA Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;
 RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
 DR EMBL; AE006574; AAK34188.1; -.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR007092; LRR_SD522.
 DR InterPro; IPR006270; Strept_his_triad.
 DR Pfam; PF00560; LRR_1_5.
 DR Pfam; PF04270; Strept_his_triad; 3.
 DR PRINTS; PR00019; LEURICHRPT.
 KW Complete proteome.
 SQ SEQUENCE 792 AA; 87458 MW; 9D5B3228485ACE0 CRC64;

Query Match 94.4%; Score 34; DB 2; Length 792;
 Best Local Similarity 85.7%; Pred. No. 2.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DPSTLKR 7
 |||||
 Db 299 DPSTIKR 305

RESULT 5

08K714 PRELIMINARY; PRT; 792 AA.

AC 08K714; Q79XH7;
 DT 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, last sequence update)
 DT 25-OCT-2004 (TREMBLrel. 28, last annotation update)
 DE Putative internalin A.
 GN Name=InLA; OrderedLocustNames=SPS0825, SPYM3_1035;
 OS Streptococcus pyogenes (serotype M3).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:15:35 ; Search time 69.1667 Seconds

(without alignments)
51.825 Million cell updates/sec

Title: US-10-646-919-33

Perfect score: 36

Sequence: 1 DPSTLKR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34	94.4	463	2	Q6BR67 debaryomyce
2	34	94.4	527	2	Q94KI0 avicennia m
3	34	94.4	792	2	Q8P0G5 streptococc
4	34	94.4	792	2	Q99276 streptococc
5	34	94.4	792	2	Q8K714 streptococc
6	33	91.7	203	2	Q9L615 arbidopsia
7	33	91.7	204	2	Q79VU7 salmoneila
8	33	91.7	204	2	Q924B8 plasmid col
9	33	91.7	204	2	Q7DJL6 plasmid r64
10	33	91.7	206	2	Q6XNAS rhodococcus
11	33	91.7	285	1	T2EB SCHPO
12	33	91.7	360	1	PLSK_DEIRA
13	33	91.7	363	2	Q8TVS1 methanopyru
14	33	91.7	540	2	Q6MDJ9 bdellovibri
15	33	91.7	1316	2	Q7Y239 oryza sativ
16	33	91.7	1316	2	Q9SL02 arbidopsia
17	33	88.9	332	2	Q6BUB5 debaryomyce
18	32	88.9	375	2	Q7RVZ7 neurospora
19	32	88.9	447	1	Y4WB_RHISN
20	32	88.9	579	2	Q8Y5P5 listeria s
21	32	88.9	579	2	Q8Y5P5 listeria mo
22	32	88.9	579	2	Q7IY10 listeria mo
23	32	88.9	654	2	Q9V5B3 dirosophila
24	32	88.9	1274	2	Q9V5B3 schizosacch
25	32	88.9	1436	2	Q7R8K3 plasmodium
26	32	88.9	2313	2	Q7S2N1 neurospora
27	32	88.9	4246	2	Q7YUR0 trypanosoma
28	31	86.1	89	2	Q9XG08 thermocoga
29	31	86.1	113	2	Q7QT74 giardia lam
30	31	86.1	162	2	Q8DX88 synchococc
31	31	86.1	221	2	Q9Y175 trichomonas

32	31	86.1	224	2	Q988D9 rhizobium 1
33	31	86.1	251	1	Y136_PSESM
34	31	86.1	253	2	Q93T30 burkholderi
35	31	86.1	253	2	Q62ES9 burkholderi
36	31	86.1	253	2	Q63YB4 burkholderi
37	31	86.1	328	1	T2EB_YEAST
38	31	86.1	343	2	Q6CEB8 saccharomyc
39	31	86.1	372	1	ARO8_PROMM
40	31	86.1	379	1	DXR_CHIMU
41	31	86.1	386	2	Q7MBW0 vibrio vuln
42	31	86.1	386	2	Q8D526 vibrio vuln
43	31	86.1	393	2	Q97107 clostridium
44	31	86.1	422	2	Q6N329 rhodospirillum
45	31	86.1	434	2	Q72R96 leptospira

ALIGNMENTS

RESULT 1	PRELIMINARY	PRT	463 AA.
Q6BR67			
AC	Q6BR67		
DT	25-OCT-2004 (TREMBLrel. 28, Created)		
DT	25-OCT-2004 (TREMBLrel. 28, Last sequence update)		
DT	25-OCT-2004 (TREMBLrel. 28, Last annotation update)		
DE	Similar to tr Q93999 Candida albicans Possible zinc-finger protein.		
GN	ORFNames=DEHA0D203179;		
OS	Debaryomyces Hansenii CBS767.		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Debaryomyces.		
OX	NCBI_TaxID=284592;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RG	STRAIN=CBS767;		
RC	Genolevures;		
RA	Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S., Lafontaine I., de Montigny J., Marck C., Neugegisse C., Talla E., Goffard N., Frangoul L., Aigle M., Anthonard V., Babour A., Barbe V., Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykassen C., Bolstrame A., Boyer J., Cattelico L., Confanlet F., de Darvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A., Hantreay F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Kerrest A., Kozul R., Lemaire M., Lesur I., Ma L., Muller H., Nicard J.M., Nikolski M., Oxtas S., Olier-Kalogeropoulos O., Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A., Swenne D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B., Zenlou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Galliard C., Weissenbach J., Wincker P., Souciet J.L.;		
RT	"Genome evolution in Yeasts."		
RL	Nature 430:35-44(2004).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RG	STRAIN=CBS767;		
RC	Genoscope;		
RA	Submitted (JUL-2004) to the EMBL/GenBank/DDA databases.		
CC	-1- SUPERCELLULAR LOCATION: Nuclear (By similarity).		
CC	-1- SIMILARITY: Contains 1 Zn(2)-Cys(6) Fungal-type binuclear cluster domain.		
DR	EMBL; CR382136; CAG87477.1; -		
DR	GO; GO:0005634; C:nucleus; IEA.		
DR	GO; GO:0003700; F:transcription factor activity; IEA.		
DR	GO; GO:0008270; F:zinc ion binding; IEA.		
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.		
DR	InterPro; IPR001138; Fungi_Tfrcp_N.		
DR	Pfam; PF00172; Zn_cfus; 1.		
DR	SMART; SM00066; GAL4; 1.		
DR	PROSITE; PS50048; ZN2_Cy6_FUNGAL_2; 1.		
KW	DNA-binding; Metal-binding; Nuclear protein; Transcription;		
KW	Transcription regulation; Zinc; Zinc-finger.		
SK	SEQUENCE 463 AA; 55028 MW; A73BB6827C2C0F73 CRC64;		

Best Local Similarity 85.7%; Pred. No. 78;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DPSTLKR 7
|||||:
Db 169 DPSTLKR 175

RESULT 3
T49600
negative acting factor related protein [imported] - *Neurospora crassa*
N:Alternate names: protein B3E4.130
C:Species: *Neurospora crassa*
C>Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49600
R:Schulte, U.; Aign, V.; Hoheisel, U.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura, submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25022
A:Accession: T49600
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-375 <SCH>
A:Cross-references: EMBL:AL355931; GSPDB:GN00116; NCSP:B3E4.130
A:Experimental source: BAC clone B3E4; strain OR74A
C:Genetics:
A:Gene: NCSP:B3E4.130
A:Map position: 6
A:Introns: 126/3

Query Match 88.9%; Score 32; DB 2; Length 375;
Best Local Similarity 85.7%; Pred. No. 35;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DPSTLKR 7
|||||:
Db 30 DPSTLKR 36

RESULT 4
AE1697
two-component sensor histidine kinase homolog lln2119 [imported] - *Listeria innocua* (str C:Species: *Listeria innocua*
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AE1697
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H. Science 294, 849-852, 2001
A:Authors: Kreft, U.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Trierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.; Title: Comparative genomics of *Listeria species*.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AE1697
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-577 <GLA>
A:Cross-references: UNIPROT:Q92A03; GB:AL592022; PIDN:CAC97349.1; PID:G16414633; GSPDB:G A:Experimental source: strain C1101262
C:Genetics:
A:Gene: lln2119

Query Match 88.9%; Score 32; DB 2; Length 577;
Best Local Similarity 85.7%; Pred. No. 55;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DPSTLKR 7
|||||:
Db 188 DPSTLKR 194

RESULT 5
AC1326
two-component sensor histidine kinase homolog lmo2011 [imported] - *Listeria monocytogenes*

C:Species: *Listeria monocytogenes*
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AC1326
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H. Science 294, 849-852, 2001
A:Authors: Kreft, U.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maok, C.; Schluter, T.; Simoes, N.; Trierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend, A.; Title: Comparative genomics of *Listeria species*.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1326
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-579 <GLA>
A:Cross-references: UNIPROT:Q81SP5; GB:NC_003210; PIDN:CAD00089.1; PID:G16411464; GSPDB: A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2011

Query Match 88.9%; Score 32; DB 2; Length 579;
Best Local Similarity 85.7%; Pred. No. 55;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DPSTLKR 7
|||||:
Db 190 DPSTLKR 196

RESULT 6
T39249
probable tripeptidylpeptidase - fission yeast (*Schizosaccharomyces pombe*)
C:Species: *Schizosaccharomyces pombe*
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T39249
R:Medler, H.; Duesterhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G. submitted to the EMBL Data Library, September 1999
A:Reference number: Z21742
A:Accession: T39249
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1274 <MED>
A:Cross-references: UNIPROT:Q9UN05; EMBL:AL117210; NID:el549906; PIDN:CAB55179.1; GSPDB: A:Experimental source: strain 972h-; clone pl p843
C:Genetics:
A:Gene: SPDB:SPAP8A3.12c
A:Map position: 1

Query Match 88.9%; Score 32; DB 2; Length 1274;
Best Local Similarity 85.7%; Pred. No. 13e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DPSTLKR 7
|||||:
Db 953 DPSTLKR 959

RESULT 7
C72286
transcription regulator, metal-sensing - *Thermotoga maritima* (strain MS8)
C:Species: *Thermotoga maritima*
C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: C72286
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M. Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: C72286
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-89 <ARN>

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:32:16 ; Search time 14.5 Seconds
(without alignments)
46.449 Million cell updates/sec

Title: US-10-646-919-33

Perfect score: 36

Sequence: 1 DPSTLKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	91.7	285	2 T41052	transcription init
2	33	91.7	1292	2 D84727	probable RAD50 DNA
3	32	88.9	375	2 T49600	negative acting fa
4	32	88.9	577	2 AE1697	two-component sens
5	32	88.9	579	2 AC1326	probable tripeptid
6	32	88.9	1274	2 T39249	transcription regu
7	31	86.1	89	2 C72286	transcription init
8	31	86.1	328	2 S18138	1-deoxy-D-xylose
9	31	86.1	379	2 G81712	probable flavoprot
10	31	86.1	393	2 B97128	xanthine dehydrog
11	31	86.1	750	2 AE2860	50S ribosomal prot
12	31	86.1	779	2 D97637	ribosomal protein
13	30	83.3	44	2 S25680	hypothetical prote
14	30	83.3	105	2 S52680	hypothetical prote
15	30	83.3	141	2 H90016	hypothetical prote
16	30	83.3	169	2 D84864	hypothetical prote
17	30	83.3	180	2 B70403	hypothetical prote
18	30	83.3	189	2 T15265	hypothetical prote
19	30	83.3	192	2 T21210	hypothetical prote
20	30	83.3	205	2 C89957	hypothetical prote
21	30	83.3	294	2 S75135	hypothetical prote
22	30	83.3	346	2 D90450	conserved hypotet
23	30	83.3	350	2 AD3560	conserved hypotet
24	30	83.3	357	2 F83195	hypothetical prote
25	30	83.3	378	2 A97469	hypothetical prote
26	30	83.3	378	2 AE2587	hypothetical prote
27	30	83.3	407	2 AG0768	probable glycosyl
28	30	83.3	407	2 G90985	hypothetical prote
29	30	83.3	407	2 B85831	hypothetical prote

30	30	83.3	407	2 F55239	hypothetical 44.9K
31	30	83.3	429	2 S67101	probable nicotinat
32	30	83.3	462	2 T36848	probable glutamine
33	30	83.3	508	2 S07546	phocoystem II chl
34	30	83.3	610	2 F88109	protein T24E12.9 l
35	30	83.3	631	2 A54659	DNA repair protein
36	30	83.3	659	2 T02838	probable membrane
37	30	83.3	712	2 AD0274	probable toxin tra
38	30	83.3	775	2 F82887	hypothetical prote
39	30	83.3	1704	2 T43141	vitelllogenin 1 - m
40	30	83.3	3938	2 T42761	Basoon protein -
41	30	83.3	3942	2 T42730	Basoon protein -
42	29	80.6	77	2 U01295	hypothetical 8.8K
43	29	80.6	99	2 E71370	probable ribosomal
44	29	80.6	196	2 AF0715	probable ABC trans
45	29	80.6	201	1 MMVWBD	24K antigen - born

ALIGNMENTS

RESULT 1
T41052
transcription initiation factor IIE beta subunit (TFIIE-BETA) - fission yeast (Schizosac
C/Species: Schizosaccharomyces pombe
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: T41052
R: Hilbert, H.; Duesterhoeft, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, August 1998
A/Reference number: Z21967
A/Accession: T41052
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-285 <HIL>
A/Cross-references: UNIPROT:P79011; EMBL:AF031324; P1DN:CAA20446.1; GSPDB:GN00068; SPDB
A/Experimental source: strain 972h-; cosmid c1672
C/Genetics:
A/Map position: 3
A/Map position: 3
C/Keywords: transcription initiation

Query Match 91.7%; Score 33; DB 2; Length 285;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7
DB 244 DPSTVXR 250

RESULT 2
D84727
probable RAD50 DNA repair protein [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C/Accession: D84727
R: Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.
eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraeier, C.M.; Venter, J.
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487; PMID:10617197
A/Accession: D84727
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-1292 <STO>
A/Cross-references: UNIPROT:Q9SL02; GB:AE002093; NID:G4263721; P1DN:AAD15407.1; GSPDB:G
C/Genetics:
A/Map position: 2
A/Map position: 2
C/Superfamily: RAD50 protein

Query Match 91.7%; Score 33; DB 2; Length 1292;

```
RESULT 2
US-10-474-792-158
; Sequence 158, Application US/10474792
; Publication No. US20040236072A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Stephen
; APPLICANT: Zagursky, Robert
; APPLICANT: Nickbarg, Elliot
; APPLICANT: Winter, Lourie
; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES
; FILE REFERENCE: AM 100399
; CURRENT APPLICATION NUMBER: US/10/474,792
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 674
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 158
; LENGTH: 343
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-474-792-158

Query Match
Best Local Similarity 94.4%; Score 34; DB 16; Length 343;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7
DB 299 DPSTLKR 305

RESULT 3
US-10-870-507A-2
; Sequence 2, Application US/10870507A
; Publication No. US20040255351A1
; GENERAL INFORMATION:
; APPLICANT: Mahajan, Pramod B.
; APPLICANT: Shi, Jinru
; TITLE OF INVENTION: Rad50 Polypeptides
; FILE REFERENCE: 1116ED
; CURRENT APPLICATION NUMBER: US/10/870,507A
; CURRENT FILING DATE: 2004-06-17
; PRIOR APPLICATION NUMBER: 60/132,575
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: 09/538,396
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1316
; TYPE: PRT
; ORGANISM: Zea mays
US-10-870-507A-2

Query Match
Best Local Similarity 91.7%; Score 33; DB 16; Length 1316;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7
DB 169 DPSTLKR 175

RESULT 4
US-10-437-963-161723
; Sequence 161723, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei

APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 161723
LENGTH: 350
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(350)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_60881C.1.pep
US-10-437-963-161723

Query Match
Best Local Similarity 88.9%; Score 32; DB 16; Length 350;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7
DB 316 DPSTLKR 322

RESULT 5
US-10-437-963-161726
; Sequence 161726, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Barbazuk, Brad
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 161726
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_60884C.1.pep
US-10-437-963-161726

Query Match
Best Local Similarity 88.9%; Score 32; DB 16; Length 351;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7
DB 317 DPSTLKR 323

RESULT 6
US-10-425-115-303434
; Sequence 303434, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comphen Ltd.

OM protein - protein search, using SW model

Run on: July 20, 2005, 13:52:31 ; Search time 71.333 Seconds
(without alignments)
38.093 Million cell updates/sec

Title: US-10-646-919-33

Perfect score: 36
Sequence: 1 DPTLTKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1736639 seqs, 388188149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US10E_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubppaa/US11A_PUBCOMB.pep:*
20: /cgn2_6/ptodata/2/pubppaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	100.0	7	18 US-10-646-919-33	Sequence 33, Appl
2	34	94.4	343	16 US-10-474-792-158	Sequence 158, App
3	33	91.7	1316	16 US-10-870-507A-2	Sequence 2, Appl1
4	32	88.9	350	16 US-10-437-963-161723	Sequence 161723,
5	32	88.9	351	16 US-10-437-963-161726	Sequence 161726,
6	32	88.9	390	16 US-10-425-115-303434	Sequence 303434,
7	32	88.9	400	15 US-10-425-114-64694	Sequence 64694, A
8	32	88.9	1156	16 US-10-794-897-6	Sequence 6, Appli
9	32	88.9	1165	13 US-10-026-188-2	Sequence 2, Appli
10	32	88.9	1274	15 US-10-369-493-2308	Sequence 2308, Ap
11	31	86.1	87	16 US-10-425-115-279521	Sequence 279521,

12	31	86.1	126	15 US-10-424-599-206235	Sequence 206235,
13	31	86.1	141	16 US-10-437-963-140157	Sequence 140157,
14	31	86.1	187	16 US-10-424-599-227418	Sequence 227418,
15	31	86.1	205	14 US-10-032-585-7467	Sequence 7467, Ap
16	31	86.1	221	17 US-10-732-923-22523	Sequence 22523, A
17	31	86.1	233	15 US-10-425-115-41550	Sequence 41550, A
18	31	86.1	269	16 US-10-425-115-19186	Sequence 19186,
19	31	86.1	328	10 US-09-893-519A-6	Sequence 6, Appl1
20	31	86.1	335	13 US-10-062-354-356	Sequence 356, App
21	31	86.1	341	16 US-10-424-599-280346	Sequence 280346,
22	31	86.1	341	16 US-10-425-115-266539	Sequence 266539,
23	31	86.1	393	15 US-10-282-122A-51738	Sequence 51738, A
24	31	86.1	495	9 US-09-738-626-6583	Sequence 6583, Ap
25	31	86.1	1410	13 US-10-014-717-3	Sequence 3, Appl1
26	31	86.1	2245	17 US-10-732-923-18446	Sequence 18446, A
27	30	83.3	7	18 US-10-646-919-34	Sequence 34, Appl
28	30	83.3	44	9 US-09-815-242-12118	Sequence 12118, A
29	30	83.3	44	15 US-10-282-122A-66859	Sequence 66859, A
30	30	83.3	53	15 US-10-424-599-187972	Sequence 187972,
31	30	83.3	58	16 US-10-425-115-184765	Sequence 184765,
32	30	83.3	83	15 US-10-424-599-173274	Sequence 173274,
33	30	83.3	141	17 US-10-470-048B-221	Sequence 221, App
34	30	83.3	178	15 US-10-425-114-67865	Sequence 67865, A
35	30	83.3	181	16 US-10-767-701-51786	Sequence 51786, A
36	30	83.3	189	14 US-10-179-766-8	Sequence 8, Appl1
37	30	83.3	201	9 US-09-815-242-5400	Sequence 5400, Ap
38	30	83.3	205	9 US-09-815-242-12561	Sequence 12561, A
39	30	83.3	205	15 US-09-815-242-13029	Sequence 13029, A
40	30	83.3	205	15 US-10-282-122A-44173	Sequence 44173, A
41	30	83.3	205	17 US-10-857-625-802	Sequence 802, App
42	30	83.3	207	15 US-10-282-122A-71415	Sequence 71415, A
43	30	83.3	241	14 US-10-156-761-14781	Sequence 14781, A
44	30	83.3	257	15 US-10-296-115-1323	Sequence 1323, Ap
45	30	83.3	297	14 US-10-166-087-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1
US-10-646-919-33
; Sequence 33, Application US/10646919
; Publication No. US20050148075A1
; GENERAL INFORMATION:
; APPLICANT: Barbas, Carlos F.
; TITLE OF INVENTION: Zinc-finger binding domains for GNN
; FILE REFERENCE: TSRI 645.2C1
; CURRENT FILING DATE: 2003-08-21
; PRIOR FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: US 09/494,190
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/EP99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthesized
US-10-646-919-33

Query Match 100.0%; Score 36; DB 18; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.6e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 DPTLTKR 7
DB 1 DPTLTKR 7

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:phage display
OTHER INFORMATION: selected and mutagenized
US-09-494-190-33

Query Match 100.0%; Score 36; DB 4; Length 7;
Best Local Similarity 100.0%; Pred. No. 4; 1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPESTLKR 7
Db 1 DPESTLKR 7

RESULT 3
US-09-645-835A-4
Sequence 4, Application US/09645835A
Patent No. 6833356
GENERAL INFORMATION:
APPLICANT: Heinrichs, Jon
APPLICANT: Johnson, Leslie S.
APPLICANT: Koenig, Scott E.
APPLICANT: Adamou, John E.
TITLE OF INVENTION: Pneumococcal Protein Homologs and Fragments for
TITLE OF INVENTION: Vaccines
FILE REFERENCE: 469201-493
CURRENT APPLICATION NUMBER: US/09/645,835A
CURRENT FILING DATE: 2000-08-25
PRIOR APPLICATION NUMBER: U.S. 60/150,750
PRIOR FILING DATE: 1999-08-25
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 792
TYPE: PRT
ORGANISM: Streptococcus pyogenes
US-09-645-835A-4

Query Match 94.4%; Score 34; DB 4; Length 792;
Best Local Similarity 85.7%; Pred. No. 58;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPESTLKR 7
Db 299 DPESTLKR 305

RESULT 4
US-09-538-396-2
Sequence 2, Application US/09538396
Patent No. 6815578
GENERAL INFORMATION:
APPLICANT: Mahajan, Pramod B.
APPLICANT: Shi, Jinrui
TITLE OF INVENTION: Maize Rad50 Orthologue and Uses Thereof
FILE REFERENCE: 1116
CURRENT APPLICATION NUMBER: US/09/538,396
CURRENT FILING DATE: 2000-03-29
EARLIER APPLICATION NUMBER: 60/132,575
EARLIER FILING DATE: 1999-05-05
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1316
TYPE: PRT
ORGANISM: Zea mays
US-09-538-396-2

Query Match 91.7%; Score 33; DB 4; Length 1316;
Best Local Similarity 85.7%; Pred. No. 1; 7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPESTLKR 7

Db 169 DPESTLKR 175

RESULT 5
US-09-248-796A-15691
Sequence 15691, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196,132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 15691
LENGTH: 345
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-15691

Query Match 88.9%; Score 32; DB 4; Length 345;
Best Local Similarity 85.7%; Pred. No. 62;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPESTLKR 7
Db 55 DPESTLKR 61

RESULT 6
US-09-248-796A-15869
Sequence 15869, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstein et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196,132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 15869
LENGTH: 210
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-15869

Query Match 86.1%; Score 31; DB 4; Length 210;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DPESTLKR 7
Db 180 DPESTLKR 186

RESULT 7
US-09-538-092-514
Sequence 514, Application US/09538092
Patent No. 6753314
GENERAL INFORMATION:
APPLICANT: Glot, Iolita
APPLICANT: Manfield, Traci A.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:17:05 ; Search time 21.6667 Seconds
(without alignments)
24.117 Million cell updates/sec

Title: US-10-646-919-33

Perfect score: 36

Sequence: 1 DPSTLKR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/prodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/iaa/DCITS_COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfillseq.pep:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	100.0	7	3	US-09-173-941-33
2	36	100.0	7	4	US-09-494-190-33
3	34	94.4	7	4	US-09-645-835A-4
4	33	91.7	1316	4	US-09-538-396-2
5	32	88.9	345	4	US-09-248-796A-15691
6	31	86.1	210	4	US-09-248-796A-15869
7	31	86.1	328	4	US-09-538-092-514
8	31	86.1	1410	3	US-09-335-409-3
9	31	86.1	1410	3	US-09-568-102-3
10	31	86.1	1410	3	US-09-567-969-3
11	31	86.1	1410	3	US-09-568-480-3
12	31	86.1	1410	3	US-09-568-486-3
13	31	86.1	1410	3	US-09-568-472-3
14	31	86.1	1410	3	US-09-567-899-3
15	30	83.3	7	3	US-09-173-941-33
16	30	83.3	7	4	US-09-494-190-33
17	30	83.3	60	4	US-09-270-767-57387
18	30	83.3	72	4	US-09-252-991A-18994
19	30	83.3	105	4	US-09-538-092-153
20	30	83.3	323	4	US-09-270-767-42120
21	30	83.3	412	4	US-09-502-540-11614
22	30	83.3	571	4	US-09-902-540-14983
23	30	83.3	830	4	US-09-252-991A-22004
24	30	83.3	1341	4	US-09-248-796A-19074
25	29	80.6	108	4	US-09-540-236-3176
26	29	80.6	201	3	US-08-369-822C-4
27	29	80.6	201	3	US-08-779-764A-20

28	29	80.6	201	3	US-08-779-764A-21	Sequence 21, Appl
29	29	80.6	201	3	US-08-779-764A-22	Sequence 22, Appl
30	29	80.6	201	3	US-08-582-776C-4	Sequence 4, Appl
31	29	80.6	201	3	US-08-434-831B-4	Sequence 4, Appl
32	29	80.6	201	4	US-09-563-456-20	Sequence 20, Appl
33	29	80.6	201	4	US-09-563-456-21	Sequence 21, Appl
34	29	80.6	201	4	US-09-563-456-22	Sequence 22, Appl
35	29	80.6	219	4	US-09-543-681A-5362	Sequence 5362, Ap
36	29	80.6	221	4	US-09-252-991A-28805	Sequence 28805, A
37	29	80.6	222	4	US-09-902-540-11510	Sequence 11510, A
38	29	80.6	296	4	US-09-270-767-41947	Sequence 41947, A
39	29	80.6	339	4	US-09-248-796A-15770	Sequence 15770, A
40	29	80.6	416	4	US-09-902-540-12512	Sequence 12512, A
41	29	80.6	696	4	US-09-134-000C-4874	Sequence 4874, Ap
42	29	80.6	1165	4	US-09-949-016-6874	Sequence 6874, Ap
43	29	80.6	1165	4	US-09-949-016-11392	Sequence 11392, A
44	28	77.8	18	6	5252328-14	Patent No. 5252328
45	28	77.8	18	6	5252328-14	Patent No. 5252328

ALIGNMENTS

```

RESULT 1
US-09-173-941-33
; Sequence 33, Application US/09173941
; Patent No. 6140081
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV00815
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
; OTHER INFORMATION: codon binding sequence
US-09-173-941-33

Query Match      100.0%; Score 36; DB 3; Length 7;
Best local Similarity 100.0%; Pred. No. 4, 1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DPSTLKR 7
      |||||
Db      1 DPSTLKR 7

RESULT 2
US-09-494-190-33
; Sequence 33, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence

```


Db 299 DPSTIKR 305

RESULT 6

ADRM3938 standard; protein; 792 AA.

ADRM3938;

02-DEC-2004 (first entry)

S. pyogenes hyperimmune system reactive antigen SpY1361 DNA.

hyperimmune serum reactive antigen; vaccine; anticaline.

Streptococcus pyogenes.

WO2004078907-A2.

16-SEP-2004.

02-MAR-2004; 2004WO-EP02087.

04-MAR-2003; 2003EP-00450061.

(INTE-) INTERCELL AG.

Weinke A, Nagy E, Winkler B, Gelbmann D;

WPI; 2004-653698/63.

N-PSDB; ADM83788.

New isolated nucleic acid molecules encoding hyperimmune serum-reactive antigens from Streptococcus pyogenes, useful for diagnosing, preventing and treating S. pyogenes infections.

Claim 14; SEQ ID NO 206; 145bp; English.

This invention describes a novel nucleic acid molecule encoding a hyperimmune serum reactive antigen or its fragment from Streptococcus pyogenes. The nucleic acid molecule or hyperimmune serum-reactive antigen or its fragment are useful for the manufacture of a pharmaceutical preparation, especially a vaccine, against S. pyogenes infection. In addition, the hyperimmune serum reactive antigen or fragment is used for the isolation and/or purification and/or identification of an interaction partner of the hyperimmune serum reactive antigen or its fragment, for the generation of a peptide (e.g. anticalines) binding to the antigen or fragment, or for the manufacture of a functional nucleic acid selected from aptamers and Spiegelmers. The nucleic acid molecule may also be used for the manufacture of functional ribonucleic acids, such as ribozymes, antisense nucleic acids and siRNA. ADM83733-ADM84189 represent S. pyogenes hyperimmune serum reactive antigens, fragments and the encoding polynucleotide described in the invention.

Sequence 792 AA;

Query Match 94.4%; Score 34; DB 8; Length 792;

Best Local Similarity 85.7%; Pred. No. 2.4e+02; Mismatches 6; Conservative 1; Indels 0; Gaps 0;

1 DPSTIKR 7

299 DPSTIKR 305

RESULT 7

ADM26711 standard; protein; 363 AA.

ADM26711;

20-MAY-2004 (first entry)

DE Hyperthermophile Methanopyrus kandleri protein #1317.

hyperthermophile; protein stability enhancement;

protein activity enhancement.

Methanopyrus kandleri.

WO2003076575-A2.

18-SEP-2003.

04-MAR-2003; 2003WO-US006664.

04-MAR-2002; 2002US-0361742P.

14-MAY-2002; 2002US-0380423P.

16-SEP-2002; 2002US-0410974P.

(FIDE-) FIDELITY SYSTEMS INC.

(MALY/) MALYKH A.

Slesarev AI, Pavlov A, Pavlova N, Kozayavkin S;

WPI; 2003-748383/70.

N-PSDB; ADM27081.

Claim 31; SEQ ID NO 1317; 1023bp; English.

The invention comprises the amino acid sequence of proteins from the hyperthermophile Methanopyrus kandleri, the invention also comprises the complete genome from Methanopyrus kandleri. The Methanopyrus kandleri proteins of the invention are useful for enhancing the stability and/or activity of other proteins. The Methanopyrus kandleri genome is useful in a variety of diagnostic and analytical methods. The present amino acid sequence represents a Methanopyrus kandleri protein of the invention.

Sequence 363 AA;

Query Match 91.7%; Score 33; DB 7; Length 363;

Best Local Similarity 85.7%; Pred. No. 1.7e+02; Mismatches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 DPSTIKR 7

137 DPSTIKR 143

RESULT 8

ABB77986 standard; protein; 1292 AA.

ABB77986;

22-OCT-2002 (first entry)

Amino acid sequence of an Arabidopsis RAD50 homologue.

Nucleic acid integration; homologous recombination; telomeric region;

RAD50.

Arabidopsis thaliana.

EP1217074-A1.

26-JUN-2002.

22-DEC-2000; 2000EP-00204693.

22-DEC-2000; 2000EP-00204693.

```

RESULT 4
AAU00027
ID AAU00027 standard; protein; 792 AA.
XX
XX AAU00027;
XX
XX 11-MAY-2001 (first entry)
XX
XX Streptococcus pyogenes GAS36(2).
XX
XX Group A streptococci protein 36(2); GAS36(2); immunogen; vaccine;
XX antibody; necrotizing fasciitis; scarlet fever; sepsis; impetigo;
XX bacterial meningitis; otitis media; community-acquired pneumonia.
XX
XX Streptococcus pyogenes.
XX
XX Key Location/Qualifiers
XX Region 94..99
XX /label= Histidine triad motif
XX /note= "Proposed to be involved in metal binding"
XX
XX Region 154..169
XX /label= Histidine triad motif
XX /note= "Proposed to be involved in metal binding"
XX
XX Region 246..251
XX /label= Histidine triad motif
XX /note= "Proposed to be involved in metal binding"
XX
XX Region 329..334
XX /label= Histidine triad motif
XX /note= "Proposed to be involved in metal binding"
XX
XX MO200114421-A1.
XX
XX 01-MAR-2001.
XX
XX 25-AUG-2000; 2000WO-US023417.
XX
XX 25-AUG-1999; 99US-0150750P.
XX
XX (MEDT-) MEDIMOTNE INC.
XX
XX Koenig S, Heinrichs J, Johnson LG, Adamou JE;
XX
XX WPI; 2001-211305/21.
XX
XX N-PSDB; AAS00037.
XX
XX New polypeptides obtained from group A or B streptococci, especially
XX Streptococcus aureus homologous to Sp36 protein of Streptococcus
XX pneumoniae useful as antibacterial vaccines.
XX
XX Claim 4; Fig 5b; 62pp; English.
XX
XX The sequence represents Streptococcus pyogenes Group A Streptococci
XX protein 36(2), GAS36(2). A recombinant cell producing GAS36, GAS36(2) or
XX GAS36 is useful as a vaccine for vaccinating an animal, preferably a
XX human against infection by a bacterial organism such as a streptococcal
XX or streptococcal bacteria, and for treating a disease caused by group A
XX streptococci, group B streptococci or Streptococcus aureus in an animal
XX preferably a human. Vaccines and antibodies against the proteins of the
XX invention are useful in prophylaxis and/or treatment of diseases such as
XX necrotizing fasciitis, scarlet fever, sepsis, impetigo, bacterial
XX meningitis, otitis media, community-acquired pneumonia and many diseases
XX of newborns. The proteins are also used as immunogens to stimulate the
XX production of antibodies for use in passive immunotherapy, for use as
XX diagnostic reagents and for use as reagents in other processes such as
XX affinity chromatography
XX
XX Sequence 792 AA;
XX
XX Query Match 94.4%; Score 34; DB 4; Length 792;
XX Best Local Similarity 85.7%; Pred. No. 2.4e+02;
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 DPTLR 7
DB 299 DPTLR 305
XX
XX RESULT 5
XX ABP27504
XX ID ABP27504 standard; protein; 792 AA.
XX
XX ABP27504;
XX
XX 02-JUL-2002 (first entry)
XX
XX Streptococcus polypeptide SEQ ID NO 4184.
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX group A streptococcus; Streptococcus pyogenes; antibacterial;
XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus pyogenes.
XX
XX MO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
XX
XX 24-NOV-2000; 2000GB-00028727.
XX
XX 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Maignani V, Margarit Y Rosl, Grandi G, Fraser C;
XX Tettein H;
XX
XX WPI; 2002-352536/38.
XX
XX N-PSDB; ABN68135.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
XX disease caused by Streptococcus bacteria, such as meningitis, and for
XX detecting a compound that binds to the protein.
XX
XX Claim 1; Page 3575; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
XX streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
XX (Streptococcus pyogenes), comprising one of 5483 sequences (s1), given in
XX the specification. The proteins have antibacterial and antiinflammatory
XX activity. (1), nucleic acids encoding (1), ABN66044-ABN71526 and
XX antibodies that bind (1) are used in the manufacture of medicaments for
XX the treatment or prevention of infection or disease caused by
XX Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX Nucleic acids encoding (1) are used to detect Streptococcus in a
XX biological sample. (1) is used to determine whether a compound binds to
XX (1). A composition comprising (1) or a nucleic acid encoding (1), may be
XX used as a vaccine or diagnostic composition. The disease caused by
XX Streptococcus that is prevented or treated may be meningitis. Nucleic
XX acid encoding (1) may be used to recombinantly produce (1) and may be
XX used in gene therapy. Antibodies to (1) are used for affinity
XX chromatography, immunoassays, and distinguishing/identifying
XX Streptococcus proteins
XX
XX Sequence 792 AA;
XX
XX Query Match 94.4%; Score 34; DB 5; Length 792;
XX Best Local Similarity 85.7%; Pred. No. 2.4e+02;
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 DPTLR 7
XX

```

CC achieved through specific contacts from side chains of amino acid
 CC residues in the alpha helix. Each zinc finger can recognise a subside of
 CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains
 CC allows the recognition of extended contiguous asymmetric DNA sequences.
 CC For example, a synthetic polyclonal protein containing six zinc finger
 CC domains can recognise an 18 bp sequence, and such proteins are
 CC potentially highly gene-specific. The novel nucleotide-binding zinc
 CC finger proteins may therefore be used in the development of artificial
 CC gene-specific transcriptional regulators. Such transcriptional switches
 CC may be used to regulate the expression of oncogenes such as erbB-2,
 CC overexpression of which is involved in malignant transformation. The
 CC proteins are therefore useful in the treatment of cancers, and may also
 CC be used to activate genes involved in fighting diseases, and to treat
 CC viral infections by inhibiting the synthesis of viral gene products. They
 CC may be used in DNA-based diagnostic applications. The proteins may also
 CC be used in producing functional gene knockout or activation in
 CC heterozygous transgenic animals. Proteins of the invention can
 CC discriminate between sequences which have a single base difference. This
 CC is manifested in a >100-fold decrease in affinity for the variant
 CC sequence. Gene activation and repression can be achieved by targeting
 CC within the gene transcript, suggesting that information obtained from
 CC expressed sequence tags may be sufficient for the construction of gene
 CC switches. Sequences AAB02876-B02869 represent zinc finger alpha helix
 CC phage library peptides disclosed in the invention
 XX

SO Sequence 7 AA:

Query Match 100.0%; Score 36; DB 3; Length 7;
 Best Local Similarity 100.0%; Pred. No. 1.8e+06;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7
 |||||
 Db 1 DPSTLKR 7

RESULT 2

ADP61982
 ID ADP61982 standard; peptide; 7 AA.

AC ADP61982;

DT 12-FEB-2004 (first entry)

XX zinc finger binding region #33.

XX zinc finger-nucleotide binding polypeptide; expression regulation;

KW zinc finger binding region.

XX Synthetic.

PN US6610512-B1.

PD 26-AUG-2003.

PF 26-JUN-2000; 2000US-00494190.

PR 16-OCT-1998; 98US-00173941.

PR 14-OCT-1999; 99WO-EP007742.

PA (SCRI) SCRIPPS RES INST.

PI Barbas CF;

XX MPI; 2003-800134/75.

PT Regulating expression of nucleotide sequence that contains sequence 5'-
 PT (GNN)n-3', comprising exposing nucleotide sequence to composition
 PT comprising isolated and purified zinc finger-nucleotide binding
 PT polypeptide.
 XX
 PS Disclosure; SEQ ID NO 33; 46pp; English.
 XX

CC The invention relates to an expression of nucleotide sequence that
 CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by
 CC exposing nucleotide sequence to composition comprising isolated and
 CC purified zinc finger-nucleotide binding polypeptide that contains a
 CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Leu-Val-
 CC Arg. The invention is used in the regulation of the expression of the
 CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present
 CC sequence represents the amino acid sequence of a zinc finger binding
 CC region.
 XX

SO Sequence 7 AA:

Query Match 100.0%; Score 36; DB 7; Length 7;
 Best Local Similarity 100.0%; Pred. No. 1.8e+06;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7
 |||||
 Db 1 DPSTLKR 7

RESULT 3

ADJ98424
 ID ADJ98424 standard; peptide; 7 AA.

XX ADJ98424;

DT 06-MAY-2004 (first entry)

DE zinc finger DNA binding peptide #95.

XX library; multimeric DNA binding polypeptide;

KW zinc finger DNA binding peptide; gene expression silencing;

XX gene expression enhancement.

XX Unidentified.

PN WO2003066828-A2.

PD 14-AUG-2003.

PF 07-FEB-2003; 2003WO-US003705.

PR 07-FEB-2002; 2002US-0354981P.

PA (SCRI) SCRIPPS RES INST.

PI Barbas CF, Blancafort P;

XX MPI; 2003-731499/69.

PT New zinc finger library of multimeric DNA binding polypeptides, useful
 PT for sterically occluding the binding site of a natural transcription
 PT factor, and enhancing or silencing target gene expression.
 XX

PS Disclosure; Fig 20; 64pp; English.

XX The invention comprises a library of multimeric DNA binding polypeptides
 CC (e.g. zinc finger DNA binding polypeptides). The zinc finger binding
 CC polypeptides of the invention are useful for sterically occluding the
 CC binding site of a natural transcription factor and enhancing or silencing
 CC target gene expression. The present amino acid sequence represents a zinc
 CC finger DNA binding peptide of the invention.
 XX

SO Sequence 7 AA:

Query Match 100.0%; Score 36; DB 7; Length 7;
 Best Local Similarity 100.0%; Pred. No. 1.8e+06;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DPSTLKR 7
 |||||
 Db 1 DPSTLKR 7

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:14:30 ; Search time 82.5 Seconds

(without alignments)
32.816 Million cell updates/sec

Title: US-10-646-919-33

Perfect score: 36

Sequence: 1 DPSTLKR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	36	100.0	7	3	AAB02892	Aab02892 Nucleotid
2	36	100.0	7	7	ADP61982	Adf61982 Zinc fing
3	36	100.0	7	7	ADJ98424	Adj98424 Zinc fing
4	34	94.4	792	4	AAU00027	Aau00027 Streptoco
5	34	94.4	792	5	ABP27504	Abp27504 Streptoco
6	34	94.4	792	8	ADR83938	Adr83938 S. pyogen
7	33	91.7	363	7	ADM26711	Adm26711 Hyperther
8	33	91.7	1292	5	AB877986	Abb77986 Amino aci
9	33	91.7	1316	4	AA827248	Aab27248 Maize RAD
10	32	88.9	579	5	AB848516	Abb48516 Listeria
11	32	88.9	651	4	AB858403	Abb58403 Drosophil
12	32	88.9	651	8	ADS96540	Ad96540 Drosophil
13	32	88.9	676	8	ADM90820	Adm90820 Human pha
14	32	88.9	928	8	ADM90818	Adm90818 Human pha
15	32	88.9	1156	8	ADR87165	Adr87165 Rat betar
16	32	88.9	1164	5	AB883853	Abb83853 Rat L-TRP
17	32	88.9	1274	8	ADN19655	Adn19655 Bacterial
18	32	88.9	1885	8	ADM90819	Adm90819 Human pha
19	31	86.1	205	5	ABP73630	Abp73630 Candida a
20	31	86.1	328	5	AAU82946	Aau82946 S. cerevi
21	31	86.1	393	6	ABU33814	Abu33814 Protein e
22	31	86.1	434	8	ADR28242	Adr28242 Leptospi
23	31	86.1	495	4	AA932829	Aag32829 C glutami
24	31	86.1	1410	3	AAV58574	Adg58574 Sorangium
25	31	86.1	1576	8	ADQ97959	Adq97959 Mouse can

26	30	83.3	7	3	AAB02893	Aab02893 Nucleotid
27	30	83.3	7	7	ADP61983	Adf61983 Zinc fing
28	30	83.3	7	7	ADJ98425	Adj98425 Zinc fing
29	30	83.3	44	4	AAU36525	Aau36525 Pseudomon
30	30	83.3	44	6	ABU38935	Abu38935 Protein e
31	30	83.3	72	7	AB070248	Ab070248 Pseudomon
32	30	83.3	105	6	ABR53589	Ab53589 Protein s
33	30	83.3	105	7	ADK64318	Adk64318 Diaseae t
34	30	83.3	141	6	ABU19003	Abu19003 Pathogen
35	30	83.3	141	6	ABM73460	Abm73460 Staphyloc
36	30	83.3	158	4	AB869105	Abb69105 Drosophil
37	30	83.3	158	8	ADS96486	Ad96486 Drosophil
38	30	83.3	169	3	AA643188	Aag43188 Arabidops
39	30	83.3	169	3	AA607957	Aag07957 Arabidops
40	30	83.3	189	6	AAO19958	Aao19958 C elegans
41	30	83.3	201	4	AAU33904	Aau33904 Staphyloc
42	30	83.3	205	4	AAU37436	Aau37436 Staphyloc
43	30	83.3	205	4	AAU36968	Aau36968 Staphyloc
44	30	83.3	205	6	ABU16249	Abu16249 Protein e
45	30	83.3	205	6	ABM71226	Abm71226 Staphyloc

ALIGNMENTS

RESULT 1
AAB02892
ID AAB02892 standard; peptide; 7 AA.
XX
AC AAB02892;
XX
DT 18-SEP-2000 (first entry)
XX
DE Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:33.
XX
KW Zinc finger domain; alpha helix; nucleotide binding; DNA binding;
KW polylactyl protein; asymmetric target recognition;
KW gene specific transcriptional regulator; gene activator; gene repressor;
KW transcriptional switch; oncogene; erbB-2; cancer; tumour; gene therapy;
KW transgenic animal; antiviral; anticancer; diagnosis.
XX
OS Synthetic.
XX
PN WO200023464-A2.
XX
PD 27-APR-2000.
XX
PF 14-OCT-1999; 99WO-EP007742.
XX
PR 16-OCT-1998; 98US-00173941.
XX
PA (NOVS) NOVARTIS AG.
PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PA (SCRT) SCRIPPS RES INST.
XX
PI Barbas CF;
XX
DR WPI, 2000-339648/29.
XX
PT Novel isolated and purified zinc finger nucleotide-binding proteins with
PT specificity for GNN triplet sequences, useful in gene therapy and for
PT regulating gene function.
XX
PS Disclosure, Fig 1; 48pp; English.
XX
CC The invention relates to zinc finger nucleotide-binding proteins which
CC comprise 2-12, preferably 2-6, operatively linked motifs selected from
CC sequences AAB02860-B02875. Sequences AAB02860-B02875 represent the alpha
CC helical regions of zinc finger domains which specifically bind to target
CC nucleotide triplets of the sequence 5'-GNN-3'. Such regions may be linked
CC by the peptide linker TGRKP (AAB02970). The Cys62-His2 zinc finger motif
CC is the most frequently utilised nucleic acid binding motif in eukaryotes,
CC and constitutes a beta-beta-alpha fold. Nucleic acid recognition is

```

DR HSP: P04050; 1150.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO: GO:0006350; P:transcription; IEA.
DR InterPro: IPR007075; RNA_pol_Rpb1_6.
DR Pfam: PF04992; RNA_pol_Rpb1_6; 1.
FT NON_TER 1 91
FT NON_TER 91 1
SQ SEQUENCE 91 AA; 10475 MW; ACA2762E06ABBEC CRC64;

Query Match
Best Local Similarity 92.1%; Score 35; DB 2; Length 91;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCRLSR 7
Db 20 GCRLSK 26

RESULT 3
O9BNP8 PRELIMINARY; PRT; 91 AA.
ID O9BNP8;
AC O9BNP8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE RNA polymerase II largest subunit (Fragment).
OS Bacthopolys multidentatus.
OC Eukaryota; Metazoa; Arthropoda; Myriapoda; Chilopoda;
OC Pleurostigmophora; Lithobiomorpha; Ethropolyidae; Bacthopolys.
OX NCBI_TaxID=118455;
RN [1]
RP SEQUENCE FROM N.A.
RA Regier J.C., Shultz J.W.
RT "A phylogenetic analysis of Myriapoda (Arthropoda) using two nuclear
   protein-encoding genes."
RL Zool. J. Linn. Soc. 132:469-486(2001).
DR EMBL: AF240897; AAK11874.1; -.
DR HSP: P04050; 1150.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO: GO:0006350; P:transcription; IEA.
DR InterPro: IPR007075; RNA_pol_Rpb1_6.
DR Pfam: PF04992; RNA_pol_Rpb1_6; 1.
FT NON_TER 1 91
FT NON_TER 91 1
SQ SEQUENCE 91 AA; 10457 MW; ACA2736B12FFBEC CRC64;

Query Match
Best Local Similarity 92.1%; Score 35; DB 2; Length 91;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCRLSR 7
Db 20 GCRLSK 26

RESULT 4
O6JSV7 PRELIMINARY; PRT; 108 AA.
ID O6JSV7;
AC O6JSV7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE RNA polymerase II largest subunit (Fragment).
OS Lithobius forticatus.
OC Eukaryota; Metazoa; Arthropoda; Myriapoda; Chilopoda;
OC Pleurostigmophora; Lithobiomorpha; Lithobiidae; Lithobius.
OX NCBI_TaxID=245740;
RN [1]
RP SEQUENCE FROM N.A.
RA Regier J.C.
RT Submitted (May-2003) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL: AY310213; AA077115.1; -.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO: GO:0006350; P:transcription; IEA.
DR InterPro: IPR007081; RNA_pol_Rpb1_5.
DR InterPro: IPR007075; RNA_pol_Rpb1_6.
DR Pfam: PF04998; RNA_pol_Rpb1_5; 1.
DR Pfam: PF04992; RNA_pol_Rpb1_6; 1.
FT NON_TER 1 108
FT NON_TER 108 1
SQ SEQUENCE 108 AA; 12041 MW; D3BD188345BFPD9E CRC64;

Query Match
Best Local Similarity 92.1%; Score 35; DB 2; Length 108;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCRLSR 7
Db 20 GCRLSK 26

RESULT 5
O823V0 PRELIMINARY; PRT; 228 AA.
ID O823V0;
AC O823V0; Q7C7B2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Probable global regulatory protein.
GN Name=kgdr; OrderedLocustNames=STRY3251, t3010;
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RA Parikh J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
   Churcher C.M., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M.,
   Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
   Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
   Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
   Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
   Quail M.A., Rutherford K.M., Simmonds W., Skelton J., Stevens K.,
   Whitehead S., Barrett B.G.
RT "Complete genome sequence of a multiple drug resistant Salmonella
   enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=Ty2 / ATCC 700931;
RC MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
RX MEDLINE=22531367; PubMed=12644504;
RX DOI=10.1128/JB.185.7.2330-2337.2003;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
   Burdand V., Kodoyianni V., Schwartz D.C., Blatner F.R.,
   "Comparative genomes of Salmonella enterica serovar Typhi strains Ty2
   and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
DR EMBL: AL627277; CAD02922.1; -.
DR EMBL: AE016844; AA070562.1; -.
KW Complete proteome.
SQ SEQUENCE 228 AA; 25191 MW; 50EA8A3B87B9D07D CRC64;

Query Match
Best Local Similarity 92.1%; Score 35; DB 2; Length 228;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCRLSR 7
Db 27 GCRLSK 33

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:15:35 ; Search time 69.1667 Seconds
(without alignments)
51.825 Million cell updates/sec

Title: US-10-646-919-32
Perfect score: 38
Sequence: 1 GCRELSR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : uniprot_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	92.1	91	2 Q6JSY5	Q6JSY5 australoblu
2	35	92.1	91	2 Q9BNK6	Q9BNK6 pokabius bi
3	35	92.1	91	2 Q9BNP8	Q9BNP8 bothroplys
4	35	92.1	108	2 Q6JSV7	Q6JSV7 lithobius f
5	35	92.1	228	2 Q8Z3V0	Q8Z3V0 salmonella
6	35	92.1	228	2 Q8ZM49	Q8ZM49 salmonella
7	34	89.5	92	2 Q00385	Q00385 coccidioid
8	34	89.5	93	2 Q7S8T6	Q7S8T6 neuropept
9	34	86.8	140	2 Q6CYK9	Q6CYK9 erwinia car
10	33	86.8	216	2 Q700H1	Q700H1 uscllago ma
11	33	86.8	216	2 Q700H8	Q700H8 uscllago ma
12	33	86.8	216	2 Q700I0	Q700I0 uscllago ma
13	33	86.8	248	2 Q7CP30	Q7CP30 agrobacteri
14	33	86.8	254	2 Q7DJ56	Q7DJ56 agrobacteri
15	33	86.8	254	2 Q9WMC0	Q9WMC0 agrobacteri
16	33	86.8	273	2 Q8VMB8	Q8VMB8 anabaena sp
17	33	86.8	410	1 B4_USITMA	B4_USITMA uscllago ma
18	33	86.8	756	2 Q8DKX1	Q8DKX1 synchococc
19	33	86.8	767	2 Q7Q6H8	Q7Q6H8 anopheles g
20	32	84.2	132	2 Q26405	Q26405 brugia mala
21	32	84.2	151	2 Q9KGF1	Q9KGF1 bacillus ha
22	32	84.2	175	2 Q7XIC9	Q7XIC9 leprospirili
23	32	84.2	225	2 Q9GTI8	Q9GTI8 wucheria
24	32	84.2	273	2 Q7SDP4	Q7SDP4 ashbya goss
25	32	84.2	282	2 Q8TXS8	Q8TXS8 methanopyru
26	32	84.2	341	2 Q8ZVY1	Q8ZVY1 pyrobaculum
27	32	84.2	352	2 Q04963	Q04963 brugia paha
28	32	84.2	380	2 Q8VNG5	Q8VNG5 mus musculu
29	32	84.2	386	2 Q6KZJ3	Q6KZJ3 picrophilus
30	32	84.2	539	2 Q9AS88	Q9AS88 oryza sativ
31	32	84.2	645	2 Q8VDM7	Q8VDM7 mus musculu

32	32	84.2	655	1 NA10_HUMAN	Q86w26 homo sapien
33	32	84.2	708	2 Q91YS3	Q91YS3 mus musculu
34	32	84.2	784	2 Q87ZN8	Q87ZN8 pseudomonas
35	32	84.2	846	2 Q8IXT0	Q8IXT0 homo sapien
36	32	84.2	853	2 Q8C1A7	Q8C1A7 mus musculu
37	32	84.2	1062	1 NA12_HUMAN	Q9nu02 homo sapien
38	32	84.2	1116	1 PLAS_HUMAN	Q9nu00 homo sapien
39	32	84.2	1205	2 Q6ZPK1	Q6ZPK1 mus musculu
40	32	84.2	1421	2 Q7XSL9	Q7XSL9 oryza sativ
41	32	84.2	2078	2 Q8WPN1	Q8WPN1 oikopleura
42	31	81.6	103	2 Q6XNB7	Q6XNB7 rhodococcus
43	31	81.6	309	1 DUSC_AZOVI	P96192 azotobacter
44	31	81.6	343	2 Q6DTJ9	Q6dt39 paenibacill
45	31	81.6	346	1 VANL_STRCO	Q9xak7 streptomyce

ALIGNMENTS

RESULT 1
Q6JSY5 PRELIMINARY; PRT; 91 AA.
ID Q6JSY5;
AC Q6JSY5;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE RNA polymerase II largeet subunit (Fragment).
OS Australobius scabrior.
OC Eukaryota; Metazoa; Arthropoda; Myriapoda; Chilopoda;
OC Pleurostigmophora; Lithodromophra; Lithodidae; Australobius.
OX NCBI_TaxID=126927;
RN (1)
RP SEQUENCE FROM N.A.
RA Regier J.C.;
RL Submitted (MAY-2003) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY310185; F:DNA binding; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR007075; RNA_pol_Rpbl_6.
DR Pfam; PF04992; RNA_pol_Rpbl_6; 1.
FT NON_TER 1
FT NON_TER 91
SQ SEQUENCE 91 AA; 10445 MW; 788C828412FFA37D CRC64;
Query Match Best Local Similarity 92.1%; Score 35; DB 2; Length 91;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCRELSR 7
Db 20 GCRELSK 26
RESULT 2
Q9BNK6 PRELIMINARY; PRT; 91 AA.
ID Q9BNK6;
AC Q9BNK6;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE RNA polymerase II largeet subunit (Fragment).
OS pokabius bilabialis.
OC Eukaryota; Metazoa; Arthropoda; Myriapoda; Chilopoda;
OC Pleurostigmophora; Lithodromophra; Lithodidae; Pokabius.
OX NCBI_TaxID=116484;
RN (1)
RP SEQUENCE FROM N.A.
RA Regier J.C.; Shultz J.W.;
RT "A phylogenetic analysis of Myriapoda (Arthropoda) using two nuclear
protein-encoding genes";
RL Zool. J. Linn. Soc. 132:469-486 (2001).
DR EMBL; AF240939; AAK11916.1; --

A:Gene: atsc
A:Genome: plasmid

Query Match

Best Local Similarity 86.8%; Score 33; DB 2; Length 248;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRLSR 7
Db 35 GCRDMSR 41

RESULT 3

hypothetical protein al15016 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120
C:Accession: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C:Accession: AH2432
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kurihara, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2432
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1273 <KUR>
A:Cross-references: UNIPROT:Q9YMB8; GB:BA000019; PIDN:BAH7615.1; PID:G17134154; GSPDB:G
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: al15016

Query Match

Best Local Similarity 86.8%; Score 33; DB 2; Length 273;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCRLSR 7
Db 243 GCRSLSR 249

RESULT 4

C32696
b4 protein - smut fungus (Ustilago maydis)

C:Species: Ustilago maydis (corn smut)
C:Date: 22-Jun-1990 #sequence_revision 28-Aug-1992 #text_change 09-Jul-2004
C:Accession: C32696
R:Schulz, B.; Banuett, F.; Dahl, M.; Schlessinger, R.; Schaefer, W.; Martin, T.; Herskowitz, Cell 60, 295-306, 1990
A:Title: The b alleles of U. maydis, whose combinations program pathogenic development.
A:Reference number: A32696; MUID:90124638; PMID:1967554
A:Accession: C32696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-409 <SCH>
A:Cross-references: UNIPROT:P32018; GB:M58556
A:Note: the authors translated the codon GAA for residue 56 as Lys
C:Superfamily: mating-type locus protein b1
C:Keywords: DNA binding; nucleus

Query Match

Best Local Similarity 86.8%; Score 33; DB 2; Length 409;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRLSR 6
Db 108 GCRELS 113

RESULT 5

B84122
hypothetical protein BH3778 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004

C:Accession: B84122

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Maui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: B84122

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <STO>

A:Cross-references: UNIPROT:Q9K6F1; GB:AP001520; GB:BA000004; NID:G10176401; PIDN:BAH074

A:Experimental source: strain C-125

C:Genetics:
A:Gene: BH3778

Query Match

Best Local Similarity 84.2%; Score 32; DB 2; Length 151;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCRLSR 7
Db 17 GCREMR 23

RESULT 6

surface-associated glycoprotein gp15/400 - nematode (Brugia pahangi) (fragment)

C:Species: Brugia pahangi

C:Date: 20-Feb-1995 #sequence_revision 23-Aug-1997 #text_change 09-Jul-2004

C:Accession: A49246; S25548

R:Tweddie, S.; Paxton, W.A.; Ingram, L.; Maizels, R.M.; McReynolds, L.A.; Selkirk, M.B.

Exp. Parasitol. 76, 156-164, 1993

A:Title: Brugia pahangi and Brugia malayi: a surface-associated glycoprotein (gp15/400)

A:Reference number: A49246; MUID:93202227; PMID:8454024

A:Accession: A49246

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-351 <TW2>

A:Cross-references: UNIPROT:Q04963; EMBL:X68190

C:Keywords: glycoprotein

Query Match

Best Local Similarity 84.2%; Score 32; DB 2; Length 351;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCRLSR 7
Db 145 GCRELR 151

RESULT 7

T35363
D-alanine-D-alanine ligase (EC 6.3.2.4) A SC66T3.06 [similarity] - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Jul-2004

C:Accession: T35363

R:Murphy, L.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, June 1999

A:Reference number: Z21576

A:Accession: T35363

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-346 <KUR>

A:Cross-references: UNIPROT:Q9XAK7; EMBL:AL079348; PIDN:CAB45462.1; GSPDB:GN00070; SCOD

A:Experimental source: strain A3(2)

C:Genetics:
A:Gene: SCODB:SC66T3.06

C:Superfamily: D-alanine-D-alanine ligase

Query Match

Best Local Similarity 81.6%; Score 31; DB 2; Length 346;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:32:16 ; Search time 14.5 seconds
(without alignments)
46.449 Million cell updates/sec

Title: US-10-646-919-32

Perfect score: 38
Sequence: 1 GCRLSR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR 79: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	92.1	228	2 AB0878	probable global re
2	33	86.8	248	2 AH3178	short chain dehydr
3	33	86.8	273	2 AH2432	hypothetical prote
4	33	86.8	409	2 C32696	b4 protein - smut
5	32	84.2	151	2 B84122	hypothetical prote
6	32	84.2	351	2 S25548	surface-associated
7	31	81.6	346	2 T35363	D-alanine-D-alanin
8	31	81.6	407	2 D69316	mRNA 3'-end proces
9	31	81.6	489	2 S44609	hypothetical prote
10	31	81.6	558	2 G83049	DNA repair protein
11	31	81.6	575	2 S43128	beta-D-glucosidase
12	31	81.6	712	2 C74149	hypothetical prote
13	31	81.6	919	2 T05746	hypothetical prote
14	30	78.9	137	2 E64845	yeast protein - Bsc
15	30	78.9	319	2 T33609	hypothetical prote
16	30	78.9	365	2 B69114	conserved hypochet
17	30	78.9	410	2 B32696	mating-type locus
18	30	78.9	420	2 I51666	Mel-1c receptor su
19	30	78.9	456	2 A31857	ribonuclease inhib
20	30	78.9	585	2 E85040	hypothetical prote
21	30	78.9	782	2 AE2262	phosphotibosylform
22	29	76.3	1607	2 T04583	TMV resistance pro
23	29	76.3	89	2 E97731	hypothetical prote
24	29	76.3	120	2 D95980	hypothetical prote
25	29	76.3	130	2 D82647	hypothetical prote
26	29	76.3	134	2 T08710	hypothetical prote
27	29	76.3	142	2 F96967	probable sigma fac
28	29	76.3	176	2 T05698	hypothetical prote
29	29	76.3	188	2 T34881	hypothetical prote

30	29	76.3	189	2 C86693	prophage pil prote
31	29	76.3	211	2 C95347	Protein (imported
32	29	76.3	270	2 S75051	lactose transport
33	29	76.3	315	2 E90435	conserved hypochet
34	29	76.3	353	2 A10220	probable sugar tra
35	29	76.3	354	2 B69284	sarcosine oxidase,
36	29	76.3	356	2 T41764	ACKMPV ORF18 - Bom
37	29	76.3	361	2 A86393	TLK7.4 protein - A
38	29	76.3	378	2 G02313	CDC37 homolog - hu
39	29	76.3	379	2 T20394	hypothetical prote
40	29	76.3	402	2 C69110	glutamate N-acetyl
41	29	76.3	417	2 A97063	NAD(PAD)-dependent
42	29	76.3	476	2 B84379	methylaspartate mu
43	29	76.3	496	2 T08674	probable finger pr
44	29	76.3	527	2 T37895	hypothetical prote
45	29	76.3	691	2 T32748	hypothetical prote

ALIGNMENTS

RESULT 1
AB0878
C:Species: *Salmonella enterica* subsp. *enterica* serovar *Typh*
A:Note: this species has also been called *Salmonella typh*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
R:Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., Connor, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N., Farrar, S., Moulie, S., O'Gaora, P.
A:Authors: Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov
A:Reference number: AB0502; MUID:21534947; PMID:11671608
A:Accession: AB0878
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-228 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD02922.1; PID:gt16504175; GSPDB:GN00176
C:Genetics:
A:gene: Kdgr

Query Match
Best Local Similarity 92.1%; Score 35; DB 2; Length 228;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRLSR 7
DB 27 GCRLSR 33

RESULT 2
AH3178
C:Species: *Agrobacterium tumefaciens*
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
A:Accession: AH3178
R:Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Chen, Y., Woo, L., erage, G., Gillet, W., Grant, C., Genthner, D., Kutyavhin, T., Levy, R., Li, M., McClell
A:Authors: Yoo, H., Teo, Y., Biddle, P., Jung, M., Kreppan, W., Perry, M., Gordon-Kamm, ster, B.W.
A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AH3178
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-248 <KUR>
A:Cross-references: UNIPROT:Q9MWC0; GB:AE008687; PIDN:AAL45846.1; PID:gt17743587; GSPDB:G
A:Experimental source: strain C58 (Dupont)
C:Genetics:


```
RESULT 2
US-10-767-701-49350
; Sequence 49350, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 49350
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3476-048-P1-K1-A4.pep
US-10-767-701-49350

Query Match      92.1%; Score 35; DB 16; Length 66;
Best Local Similarity 85.7%; Pred. No. 42;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCRCRSR 7
      |||||
      40 GCRCRSR 46

RESULT 3
US-10-029-386-31910
; Sequence 31910, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31910
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004495.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.91
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.61
US-10-029-386-31910

Query Match      86.8%; Score 33; DB 14; Length 33;
Best Local Similarity 85.7%; Pred. No. 53;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GCRCRSR 7
      |||||
      26 GCRCRSR 32

RESULT 4
US-10-425-115-269178
; Sequence 269178, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
```

```
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 269178
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_177089C.1.pep
US-10-425-115-269178

Query Match      86.8%; Score 33; DB 16; Length 76;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCRCRS 6
      |||||
      6 GCRCRS 11

RESULT 5
US-10-425-115-258881
; Sequence 258881, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 258881
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_167686C.1.pep
US-10-425-115-258881

Query Match      86.8%; Score 33; DB 16; Length 90;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GCRCRS 6
      |||||
      84 GCRCRS 89

RESULT 6
US-10-106-698-4323
; Sequence 4323, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruden et al.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:52:31 ; Search time 71.333 Seconds
(without alignments)
38.093 Million cell updates/sec

Title: US-10-646-919-32

Perfect score: 38

Sequence: 1 GCRESLR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1736639 seqs, 388188149 residues

Total number of hits satisfying chosen parameters: 1736639

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10D_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10E_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US10F_PUBCOMB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US10G_PUBCOMB.pep:*
19: /cgn2_6/ptodata/2/pubppaa/US11_PUBCOMB.pep:*
20: /cgn2_6/ptodata/2/pubppaa/US11_NEW_PUB.pep:*
21: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
22: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	7	18	US-10-646-919-32
2	35	92.1	66	16	US-10-767-701-48350
3	33	86.8	33	14	US-10-029-386-31910
4	33	86.8	76	16	US-10-425-115-269178
5	33	86.8	90	16	US-10-425-115-258881
6	32	84.2	103	14	US-10-106-698-4323
7	32	84.2	131	15	US-10-437-963-145479
8	32	84.2	147	16	US-10-437-963-145479
9	32	84.2	173	16	US-10-767-701-39771
10	32	84.2	199	16	US-10-425-115-286717
11	32	84.2	317	15	US-10-407-866-113

12	32	84.2	412	15	US-10-094-749-2350	Sequence 2350, Ap
13	32	84.2	418	16	US-10-425-115-264314	Sequence 264314,
14	32	84.2	422	15	US-10-425-115-56584	Sequence 56584, A
15	32	84.2	504	10	US-09-965-621-22	Sequence 22, Appl
16	32	84.2	504	15	US-10-407-866-22	Sequence 22, Appl
17	32	84.2	504	16	US-10-781-294-22	Sequence 22, Appl
18	32	84.2	555	18	US-10-499-353A-655	Sequence 655, App
19	32	84.2	567	16	US-10-437-963-106841	Sequence 106841,
20	32	84.2	590	15	US-10-407-866-94	Sequence 94, Appl
21	32	84.2	612	16	US-10-437-963-186933	Sequence 186933,
22	32	84.2	650	15	US-10-407-866-85	Sequence 85, Appl
23	32	84.2	650	15	US-10-407-866-86	Sequence 86, Appl
24	32	84.2	655	14	US-10-124-498-13	Sequence 13, Appl
25	32	84.2	655	14	US-10-066-521-13	Sequence 2, Appl1
26	32	84.2	655	14	US-10-296-539-2	Sequence 2, Appl1
27	32	84.2	655	15	US-10-407-866-84	Sequence 84, Appl
28	32	84.2	655	16	US-10-794-342-15	Sequence 15, Appl
29	32	84.2	770	16	US-10-343-663A-68	Sequence 68, Appl
30	32	84.2	1033	13	US-10-127-516-2	Sequence 2, Appl1
31	32	84.2	1033	13	US-10-027-629-2	Sequence 2, Appl1
32	32	84.2	1033	14	US-10-029-347-4	Sequence 4, Appl1
33	32	84.2	1033	14	US-10-132-967-2	Sequence 2, Appl1
34	32	84.2	1033	16	US-10-882-761-4	Sequence 4, Appl1
35	32	84.2	1062	14	US-10-233-663-43	Sequence 43, Appl
36	32	84.2	1094	15	US-10-275-107-60	Sequence 60, Appl
37	32	84.2	1116	9	US-09-790-318-2	Sequence 2, Appl1
38	32	84.2	1116	15	US-10-112-944-323	Sequence 323, App
39	32	84.2	1121	16	US-10-343-663A-27	Sequence 27, Appl
40	32	84.2	1421	16	US-10-437-963-182652	Sequence 182652,
41	31	81.6	43	16	US-10-425-115-25124	Sequence 215124,
42	31	81.6	70	15	US-10-424-599-149899	Sequence 149899,
43	31	81.6	70	16	US-10-425-115-313580	Sequence 313580,
44	31	81.6	82	16	US-10-425-115-283149	Sequence 283149,
45	31	81.6	104	16	US-10-425-115-196784	Sequence 196784,

ALIGNMENTS

RESULT 1
US-10-646-919-32
; Sequence 32, Application US/10646919
; Publication No. US20050148075A1
; GENERAL INFORMATION:
; APPLICANT: Barbas, Carlos F.
; TITLE OF INVENTION: Zinc-finger binding domains for GNN
; FILE REFERENCE: TSRI 645.2C1
; CURRENT APPLICATION NUMBER: US/10/646,919
; CURRENT FILING DATE: 2003-08-21
; PRIOR APPLICATION NUMBER: US 09/494,190
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: PCT/EP99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthesized
US-10-646-919-32

Query Match 100.0%; Score 38; DB 18; Length 7;
Best Local Similarity 100.0%; Pred. No. 1.6e+06;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCRESLR 7
DB 1 GCRESLR 7

```
FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:phage display
; OTHER INFORMATION: selected and mutagenized
US-09-494-190-32
```

```
Query Match
Best Local Similarity 100.0%; Score 38; DB 4; Length 7;
Pred. No. 4.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GCRELSR 7
|||:|
Db 1 GCRELSR 7
```

```
RESULT 3
US-09-270-767-33518
; Sequence 33518, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 33518
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-33518
```

```
Query Match
Best Local Similarity 81.6%; Score 31; DB 4; Length 153;
Pred. No. 1.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GCRELSR 7
|||:|
Db 84 GCRELSQ 90
```

```
RESULT 4
US-09-270-767-48735
; Sequence 48735, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 48735
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-48735
```

```
Query Match
Best Local Similarity 81.6%; Score 31; DB 4; Length 153;
Pred. No. 1.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GCRELSR 7
|||:|
Db 84 GCRELSQ 90
```

```
RESULT 5
US-09-252-991A-30735
; Sequence 30735, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
```

```
APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30735
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30735
```

```
Query Match
Best Local Similarity 81.6%; Score 31; DB 4; Length 425;
Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GCRELSR 7
|||:|
Db 287 GCRDLGR 293
```

```
RESULT 6
US-09-252-991A-27035
; Sequence 27035, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27035
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27035
```

```
Query Match
Best Local Similarity 81.6%; Score 31; DB 4; Length 708;
Pred. No. 4.5e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GCRELSR 7
|||:|
Db 301 GCOELAR 307
```

```
RESULT 7
US-09-621-976-7590
; Sequence 7590, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Joubert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent. pm
; SEQ ID NO 7590
; LENGTH: 60
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:17:05 ; Search time 21.6667 Seconds
(without alignments)
24.117 Million cell updates/sec

Title: US-10-646-919-32

Perfect score: 38
Sequence: 1 GCRELSR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/1/iaa/5A-COMB.pep:*
2: /cgn2_6/prodata/1/iaa/5B-COMB.pep:*
3: /cgn2_6/prodata/1/iaa/6A-COMB.pep:*
4: /cgn2_6/prodata/1/iaa/6B-COMB.pep:*
5: /cgn2_6/prodata/1/iaa/PCITUS-COMB.pep:*
6: /cgn2_6/prodata/1/iaa/backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	100.0	7	US-09-173-941-32	Sequence 32, Appl
2	38	100.0	7	US-09-494-190-32	Sequence 32, Appl
3	31	81.6	153	US-09-270-767-33518	Sequence 33518, A
4	31	81.6	153	US-09-270-767-48735	Sequence 48735, A
5	31	81.6	425	US-09-252-991A-30735	Sequence 30735, A
6	31	81.6	708	US-09-252-991A-27035	Sequence 27035, A
7	30	78.9	60	US-09-621-976-7550	Sequence 7550, Ap
8	30	78.9	177	US-09-270-767-46995	Sequence 46995, A
9	30	78.9	247	US-10-101-464A-730	Sequence 730, App
10	30	78.9	303	US-09-252-991A-23888	Sequence 23888, A
11	30	78.9	397	US-09-489-039A-9401	Sequence 9401, Ap
12	30	78.9	412	US-09-252-991A-26284	Sequence 26284, A
13	30	78.9	420	US-08-466-103A-2	Sequence 2, Appl1
14	30	78.9	456	US-08-910-731-2	Sequence 2, Appl1
15	30	78.9	456	US-08-795-395-2	Sequence 51, Appl
16	30	78.9	503	US-09-823-038A-51	Sequence 42310, A
17	30	78.9	648	US-09-270-767-42310	Sequence 7729, Ap
18	29	76.3	73	US-09-489-039A-7729	Sequence 5991, Ap
19	29	76.3	87	US-09-513-999C-5591	Sequence 5990, Ap
20	29	76.3	120	US-09-513-999C-5591	Sequence 23882, A
21	29	76.3	122	US-09-248-796A-23882	Sequence 46, Appl
22	29	76.3	143	US-08-253-155A-46	Sequence 11698, A
23	29	76.3	145	US-09-489-039A-11698	Sequence 20088, A
24	29	76.3	149	US-09-252-991A-20088	Sequence 424, App
25	29	76.3	179	US-09-205-258-424	Sequence 47629, A
26	29	76.3	193	US-09-270-767-47629	Sequence 23253, A
27	29	76.3	201	US-09-252-991A-23253	

28	29	76.3	206	1	US-08-468-847B-2	Sequence 2, Appl1
29	29	76.3	206	1	US-08-468-847B-20	Sequence 20, Appl
30	29	76.3	206	3	US-09-188-930-159	Sequence 159, App
31	29	76.3	206	3	US-09-188-930-266	Sequence 286, App
32	29	76.3	206	4	US-09-312-283C-159	Sequence 159, App
33	29	76.3	206	4	US-09-312-283C-286	Sequence 286, App
34	29	76.3	208	4	US-09-902-540-11472	Sequence 11472, A
35	29	76.3	212	4	US-09-902-540-12720	Sequence 12720, A
36	29	76.3	218	3	US-08-675-885-5	Sequence 5, Appl1
37	29	76.3	223	4	US-09-684-405-13	Sequence 13, Appl
38	29	76.3	244	4	US-09-902-540-10220	Sequence 10220, A
39	29	76.3	246	4	US-09-252-991A-23684	Sequence 23684, A
40	29	76.3	271	4	US-09-540-236-2146	Sequence 2146, Ap
41	29	76.3	279	4	US-09-252-991A-24954	Sequence 24954, A
42	29	76.3	281	4	US-09-252-991A-23775	Sequence 23775, A
43	29	76.3	293	4	US-10-237-551-211	Sequence 211, App
44	29	76.3	310	4	US-09-543-681A-5247	Sequence 5247, Ap
45	29	76.3	313	4	US-09-252-991A-25230	Sequence 25230, A

ALIGNMENTS

```

RESULT 1
; Sequence 32, Application US/09173941
; Patent No. 6140681
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: NOV00815
; CURRENT APPLICATION NUMBER: US/09/173,941
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: nucleotide
; OTHER INFORMATION: codon binding sequence
US-09-173-941-32

Query Match      100.0%; Score 38; DB 3; Length 7;
Best Local Similarity 100.0%; Pred. No. 4.1e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0;

QY      1 GCRELSR 7
      |||||
Db      1 GCRELSR 7

RESULT 2
US-09-494-190-32
; Sequence 32, Application US/09494190
; Patent No. 6610512
; GENERAL INFORMATION:
; APPLICANT: BARBAS, Carlos F.
; TITLE OF INVENTION: ZINC FINGER BINDING DOMAINS FOR GNN
; FILE REFERENCE: TSRI 645.2
; CURRENT APPLICATION NUMBER: US/09/494,190
; CURRENT FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: EP/99/07742
; PRIOR FILING DATE: 1999-10-14
; PRIOR APPLICATION NUMBER: US 09/173,941
; PRIOR FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 7
; TYPE: PRT
; ORGANISM: Artificial Sequence

```

QY 1 GCRELSR 7
 XX |||||
 DB 52 GCRELSR 58

RESULT 6
 AAM06550
 ID AAM06550 standard; protein; 74 AA.

XX AAM06550;

DT 05-OCT-2001 (first entry)

DE Human foetal protein, SEQ ID NO: 281.

XX Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
 KW neurotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
 KW gene therapy; antisense therapy; cancer; immune disorder;
 KW growth disorder; osteoporosis; thrombolytic disorder;
 KW nervous system disorder; inflammation.

OS Homo sapiens.

PN WO200155339-A2.

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US002723.

PR 25-JAN-2000; 2000US-00491404.

PR 15-SEP-2000; 2000US-00663870.

PR 06-NOV-2000; 2000US-00707351.

PA (HYSE-) HYSEQ INC.

PI Yeung G, Ford JB, Boyle BJ, Arterburn MC, Dmanac RA, Tang YT;
 PI Liu C, Asundi V, Zhou P, Werhman T;

DR WPI; 2001-465571/50.

DR N-PSDB; AAH94225.

PT Novel fetal proteins useful for the treatment and diagnosis of diseases
 PT associated with dysfunction of the protein e.g. cancers, immune
 PT disorders, growth disorders, thrombolytic disorders, nervous system
 PT disorders and inflammation.

PS Claim 10; Page 282-283; 715pp; English.

CC The invention relates to novel foetal polypeptides encoded by
 CC polynucleotides comprising one of 477 sequences fully defined in the
 CC specification. The foetal polynucleotides and polypeptides are useful in
 CC the treatment and diagnosis of diseases such as cancers, immune
 CC disorders, growth disorders (e.g. osteoporosis), thrombolytic disorders,
 CC nervous system disorders and inflammation. The present sequence is a
 CC polypeptide encoded by a cDNA assembled using an expressed sequence tag
 CC (EST) found to be expressed in human foetal tissue cDNA libraries

XX Sequence 74 AA;

QY Query Match 89.5%; Score 34; DB 4; Length 74;
 XX Best Local Similarity 85.7%; Pred. No. 64;
 XX Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1 GCRELSR 7
 XX |||||
 DB 44 GCREFSR 50

RESULT 7
 AAM06727
 ID AAM06727 standard; protein; 74 AA.

AC AAM06727;
 XX 05-OCT-2001 (first entry)
 DT |||||
 DE Human foetal protein, SEQ ID NO: 458.

XX Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
 KW neurotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
 KW gene therapy; antisense therapy; cancer; immune disorder;
 KW growth disorder; osteoporosis; thrombolytic disorder;
 KW nervous system disorder; inflammation.

OS Homo sapiens.

PN WO200155339-A2.

PD 02-AUG-2001.

PF 25-JAN-2001; 2001WO-US002723.

PR 25-JAN-2000; 2000US-00491404.

PR 15-SEP-2000; 2000US-00663870.

PR 06-NOV-2000; 2000US-00707351.

PA (HYSE-) HYSEQ INC.

PI Yeung G, Ford JB, Boyle BJ, Arterburn MC, Dmanac RA, Tang YT;
 PI Liu C, Asundi V, Zhou P, Werhman T;

DR WPI; 2001-465571/50.

DR N-PSDB; AAH94402.

PT Novel fetal proteins useful for the treatment and diagnosis of diseases
 PT associated with dysfunction of the protein e.g. cancers, immune
 PT disorders, growth disorders, thrombolytic disorders, nervous system
 PT disorders and inflammation.

PS Claim 10; Page 345; 715pp; English.

CC The invention relates to novel foetal polypeptides encoded by
 CC polynucleotides comprising one of 477 sequences fully defined in the
 CC specification. The foetal polynucleotides and polypeptides are useful in
 CC the treatment and diagnosis of diseases such as cancers, immune
 CC disorders, growth disorders (e.g. osteoporosis), thrombolytic disorders,
 CC nervous system disorders and inflammation. The present sequence is a
 CC polypeptide encoded by a cDNA assembled using an expressed sequence tag
 CC (EST) found to be expressed in human foetal tissue cDNA libraries

XX Sequence 74 AA;

QY Query Match 89.5%; Score 34; DB 4; Length 74;
 XX Best Local Similarity 85.7%; Pred. No. 64;
 XX Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

DB 1 GCRELSR 7
 XX |||||
 DB 44 GCREFSR 50

RESULT 8
 AAM06747
 ID AAM06747 standard; protein; 74 AA.

XX AAM06747;

DT 05-OCT-2001 (first entry)

DE Human foetal protein, SEQ ID NO: 955.

XX Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;
 KW neurotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;
 KW gene therapy; antisense therapy; cancer; immune disorder;
 KW growth disorder; osteoporosis; thrombolytic disorder;

RESULT 4
AAU53152
ID AAU53152 standard; protein; 65 AA.
XX
AC AAU53152;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #14048.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neutroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012865.
XX
PR 21-APR-2000; 2000US-019047P.
XX
PR 02-JUN-2000; 2000US-0208841P.
XX
PR 07-JUL-2000; 2000US-0216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR MPI: 2001-616774/71.
XX
DR N-PSDB; AAS59558.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
XX
XX Example 1; SEQ ID NO 14347; 1069bp; English.
XX
CC Sequences AAU3105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 65 AA;

Query Match 89.5%; Score 34; DB 4; Length 65;
Best Local Similarity 85.7%; Pred. No. 57;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 5
ABM49671
ID ABM49671 standard; protein; 65 AA.
XX
AC ABM49671;
XX
DT 20-OCT-2003 (first entry)
XX
DE Propionibacterium acnes predicted ORF-encoded polypeptide #14347.
XX
KW Acne vulgaris; antisephorhoic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine.
XX
OS Propionibacterium acnes.
XX
PN WO2003033515-A1.
XX
PD 24-APR-2003.
XX
PF 11-OCT-2002; 2002WO-US032727.
XX
PR 15-OCT-2001; 2001US-00978825.
XX
PA (CORI-) CORIXA CORP.
XX
PI Mitcham JL, Skeiky YAM, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallieve-Dougllass J;
XX
DR MPI: 2003-381789/36.
XX
DR N-PSDB; ACF64487.
XX
PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
XX
XX Example 1; SEQ ID NO 14347; 1481bp; English.
XX
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polynucleotides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridization. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a polypeptide predicted to be encoded by an ORF (open
CC reading frame) contained within the P. acnes polynucleotides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 65 AA;

Query Match 89.5%; Score 34; DB 6; Length 65;
Best Local Similarity 85.7%; Pred. No. 57;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

CC achieved through specific contacts from side chains of amino acid
 CC residues in the alpha helix. Each zinc finger can recognise a substrate of
 CC 3 bp in target DNA. Covalent linkage of multiple zinc finger domains
 CC allows the recognition of extended contiguous asymmetric DNA sequences.
 CC For example, a synthetic poly(acyl) protein containing six zinc finger
 CC domains can recognise an 18 bp sequence, and such proteins are
 CC potentially highly gene-specific. The novel nucleotide-binding zinc
 CC finger proteins may therefore be used in the development of artificial
 CC gene-specific transcriptional regulators. Such transcriptional switches
 CC may be used to regulate the expression of oncogenes such as *c-myc*,
 CC overexpression of which is involved in malignant transformation. The
 CC proteins are therefore useful in the treatment of cancers, and may also
 CC be used to activate genes involved in fighting diseases, and to treat
 CC viral infections by inhibiting the synthesis of viral gene products. They
 CC may be used in DNA-based diagnostic applications. The proteins may also
 CC be used in producing functional gene knockout or activation in
 CC heterozygous transgenic animals. Proteins of the invention can
 CC discriminate between sequences which have a single base difference. This
 CC is manifested in a >100-fold decrease in affinity for the variant
 CC sequence. Gene activation and repression can be achieved by targeting
 CC within the gene transcript, suggesting that information obtained from
 CC expressed sequence tags may be sufficient for the construction of gene
 CC switches. Sequences AAB8287-802869 represent zinc finger alpha helix
 CC phage library peptides disclosed in the invention
 CC
 CC Sequence 7 AA:

Query Match	100.0%;	Score 38;	DB 3;	Length 7;
Best Local Similarity	100.0%;	Pred. No. 1.8e+06;		
Matches	7;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

QY	1	GCRELSR	7
Db	1	GCRELSR	7

RESULT 2
ADf61981
ID ADf61981 standard; peptide; 7 AA.

AC	ADF61981;
XX	
DT	12-FEB-2004 (first entry)

DE Zinc finger binding region #32.

KW zinc finger-nucleotide binding polypeptide; expression regulation, zinc finger binding region.

OS Synthetic.

PN US6610512-B1

PD 26-AUG-2003.

PF 28-JAN-2000; 2000US-00494190.

PR 16-OCT-1998; 98US-00173941.

XX

XX

33 32 31 30 29 28 27 26 25 24 23 22 21 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

33

(GNN) n-3', comprises exposing nucleotide sequence to composition

PT polypeptide.

PS Claim 1; SEQ ID NO 32; 46pp; English
vxx

CC The invention relates to an expression of nucleotide sequence that
CC contains sequence 5'-(GNN)n-3', in which n is 1-6, is regulated by
CC exposing nucleotide sequence to composition comprising isolated and
CC purified zinc finger-nucleotide binding polypeptide that contains a
CC nucleotide binding region having a sequence of Gly-Cys-Arg-Glu-Leu-Val-
CC Arg. The invention is used in the regulation of the expression of the
CC nucleotide sequence that contains sequence 5'-(GNN)n-3'. The present
CC sequence represents the amino acid sequence of a zinc finger binding
CC region.
xx
xx Sequence 7 AA:
50

Query Match	100.0%;	Score 38;	DB 7;	Length 7;
Best Local Similarity	100.0%;	Pred. No. 1.8e+06;		
Matches	7;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

Qy	1	GCRELSR	7
Db	1	GCRELSR	7

```

RESULT 3
ADJ98423
ID      ADJ98423  standard; peptide; 7 AA.

```

06-MAY-2004 (first entry)
zinc finger DNA binding peptide #94

KM library; multimeric DNA binding polypeptide;
KM zinc finger DNA binding peptide; gene expression silencing;
KM gene expression enhancement.

Unidentified.

PN WO2003066828-A2

PD 14-AUG-2003

PF 07-FEB-2003; 2003WO-US003705.

PR 07-FEB-2002; 2002US-0354981P.

PA (SCRI) SCRIPPS RES INST.

PI Barbas CF, Blancafort P;

DR WPI; 2003-731499/69.

PT New zinc finger library of multimeric DNA binding polypeptides, useful

PT factor, and enhancing or silencing target gene expression

PS Disclosure; Fig 20; 64pp; English.

CC The invention comprises a library of multimeric DNA binding polypeptides

CC polypeptides of the invention are useful for sterically occluding the

CC target gene expression. The present amino acid sequence represents a zinc

XX

1. **Introduction**

2. **Background**

3. **Methods**

4. **Results**

5. **Conclusion**

6. **References**

7. **Appendix**

8. **Table 1**

9. **Table 2**

10. **Table 3**

11. **Table 4**

12. **Table 5**

13. **Table 6**

14. **Table 7**

15. **Table 8**

16. **Table 9**

17. **Table 10**

18. **Table 11**

19. **Table 12**

20. **Table 13**

21. **Table 14**

22. **Table 15**

23. **Table 16**

24. **Table 17**

25. **Table 18**

26. **Table 19**

27. **Table 20**

28. **Table 21**

29. **Table 22**

30. **Table 23**

31. **Table 24**

32. **Table 25**

33. **Table 26**

34. **Table 27**

35. **Table 28**

36. **Table 29**

37. **Table 30**

38. **Table 31**

39. **Table 32**

40. **Table 33**

41. **Table 34**

42. **Table 35**

43. **Table 36**

44. **Table 37**

45. **Table 38**

46. **Table 39**

47. **Table 40**

48. **Table 41**

49. **Table 42**

50. **Table 43**

51. **Table 44**

52. **Table 45**

53. **Table 46**

54. **Table 47**

55. **Table 48**

56. **Table 49**

57. **Table 50**

58. **Table 51**

59. **Table 52**

60. **Table 53**

61. **Table 54**

62. **Table 55**

63. **Table 56**

64. **Table 57**

65. **Table 58**

66. **Table 59**

67. **Table 60**

68. **Table 61**

69. **Table 62**

70. **Table 63**

71. **Table 64**

72. **Table 65**

73. **Table 66**

74. **Table 67**

75. **Table 68**

76. **Table 69**

77. **Table 70**

78. **Table 71**

79. **Table 72**

80. **Table 73**

81. **Table 74**

82. **Table 75**

83. **Table 76**

84. **Table 77**

85. **Table 78**

86. **Table 79**

87. **Table 80**

88. **Table 81**

89. **Table 82**

90. **Table 83**

91. **Table 84**

92. **Table 85**

93. **Table 86**

94. **Table 87**

95. **Table 88**

96. **Table 89**

97. **Table 90**

98. **Table 91**

99. **Table 92**

100. **Table 93**

101. **Table 94**

102. **Table 95**

103. **Table 96**

104. **Table 97**

105. **Table 98**

106. **Table 99**

107. **Table 100**

108. **Table 101**

109. **Table 102**

110. **Table 103**

111. **Table 104**

112. **Table 105**

113. **Table 106**

114. **Table 107**

115. **Table 108**

116. **Table 109**

117. **Table 110**

118. **Table 111**

119. **Table 112**

120. **Table 113**

121. **Table 114**

122. **Table 115**

123. **Table 116**

124. **Table 117**

125. **Table 118**

126. **Table 119**

127. **Table 120**

128. **Table 121**

129. **Table 122**

130. **Table 123**

131. **Table 124**

132. **Table 125**

133. **Table 126**

134. **Table 127**

135. **Table 128**

136. **Table 129**

137. **Table 130**

138. **Table 131**

139. **Table 132**

140. **Table 133**

141. **Table 134**

142. **Table 135**

143. **Table 136**

144. **Table 137**

145. **Table 138**

146. **Table 139**

147. **Table 140**

148. **Table 141**

149. **Table 142**

150. **Table 143**

151. **Table 144**

152. **Table 145**

153. **Table 146**

154. **Table 147**

155. **Table 148**

156. **Table 149**

157. **Table 150**

158. **Table 151**

159. **Table 152**

160. **Table 153**

161. **Table 154**

162. **Table 155**

163. **Table 156**

164. **Table 157**

165. **Table 158**

166. **Table 159**

167. **Table 160**

168. **Table 161**

169. **Table 162**

170. **Table 163**

171. **Table 164**

172. **Table 165**

173. **Table 166**

174. **Table 167**

175. **Table 168**

176. **Table 169**

177. **Table 170**

178. **Table 171**

179. **Table 172**

180. **Table 173**

181. **Table 174**

182. **Table 175**

183. **Table 176**

184. **Table 177**

185. **Table 178**

186. **Table 179**

187. **Table 180**

188. **Table 181**

189. **Table 182**

190. **Table 183**

191. **Table 184**

192. **Table 185**

193. **Table 186**

194. **Table 187**

195. **Table 188**

196. **Table 189**

197. **Table 190**

198. **Table 191**

199. **Table 192**

200. **Table 193**

201. **Table 194**

202. **Table 195**

203. **Table 196**

204. **Table 197**

205. **Table 198**

206. **Table 199**

207. **Table 200**

208. **Table 201**

209. **Table 202**

210. **Table 203**

211. **Table 204**

212. **Table 205**

213. **Table 206**

214. **Table 207**

215. **Table 208**

216. **Table 209**

217. **Table 210**

218. **Table 211**

219. **Table 212**

220. **Table 213**

221. **Table 214**

222. **Table 215**

223. **Table 216**

224. **Table 217**

225. **Table 218**

226. **Table 219**

227. **Table 220**

228. **Table 221**

229. <

Best Local Similarity 100.0%; Pred. No. 1.8e+06;

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

2
3
4
5
6
7
8
9

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 20, 2005, 13:14:30 ; Search time 82.5 Seconds
(without alignments)
32.816 Million cell updates/sec

Title: US-10-646-919-32

Perfect score: 38
Sequence: 1 GCRELSR 7

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	38	100.0	7	3	AAB02891	Aab02891 Nucleotid
2	38	100.0	7	7	ADP61981	Adf61981 Zinc fing
3	38	100.0	7	7	ADJ98423	Adj98423 Zinc fing
4	34	89.5	65	4	AAU53152	Aau53152 Protonib
5	34	89.5	65	6	ABM49671	Abm49671 Protonib
6	34	89.5	74	4	AAK06550	Aam06550 Human foe
7	34	89.5	74	4	AAK06727	Aam06727 Human foe
8	34	89.5	74	4	AAK06747	Aam06747 Human foe
9	34	89.5	211	4	ABG02364	Abg02364 Novel hum
10	33	86.8	33	8	ABO58276	Abos8276 Human gen
11	33	86.8	139	4	AAU50470	Aau50470 Protonib
12	33	86.8	139	6	ABM46989	Abm46989 Protonib
13	32	84.2	103	4	AAK73549	Aag73549 Human col
14	32	84.2	132	8	ADQ93912	Adq93912 Brugia ma
15	32	84.2	175	4	ABG28682	Abg28682 Novel hum
16	32	84.2	196	6	ABM65948	Abm65948 Protonib
17	32	84.2	219	4	AAU48517	Aau48517 Protonib
18	32	84.2	219	6	ABM45036	Abm45036 Protonib
19	32	84.2	235	8	ADQ93911	Adq93911 Hytherer
20	32	84.2	282	7	ADM25976	Adm25976 Hyperther
21	32	84.2	297	4	ABG21264	Abg21264 Novel hum
22	32	84.2	317	8	ADJ19422	Adj19422 Human PAN
23	32	84.2	341	4	ABG02678	Abg02678 Novel hum
24	32	84.2	352	8	ADQ93905	Adq93905 Brugia pa
25	32	84.2	352	8	ADQ93919	Adq93919 Brugia pa

26	32	84.2	412	6	ADA54782	Ada54782 Human pro
27	32	84.2	438	4	AAB94450	Aab94450 Human pro
28	32	84.2	466	8	ADP47893	Adp47893 Human CAT
29	32	84.2	470	8	ADP47892	Adp47892 Human CAT
30	32	84.2	504	5	AAO17865	Aao17865 Pyrin dom
31	32	84.2	504	5	ADG36415	Adg36415 Human PAA
32	32	84.2	504	8	ADJ19337	Adj19337 Human PAN
33	32	84.2	519	4	ABG14808	Abg14808 Novel hum
34	32	84.2	555	7	ADP14024	Adp14024 Human end
35	32	84.2	590	8	ADJ19403	Adj19403 Human PAN
36	32	84.2	609	4	ABG28217	Abg28217 Novel hum
37	32	84.2	650	8	ADJ19394	Adj19394 Human PJJ
38	32	84.2	650	8	ADJ19395	Adj19395 Human PAN
39	32	84.2	655	5	AAO15588	Aao15588 Human PYR
40	32	84.2	655	5	ABB07264	Abb07264 Human APR
41	32	84.2	655	7	ADP94751	Adp94751 Human PYR
42	32	84.2	655	8	ADJ19393	Adj19393 Human PAN
43	32	84.2	753	7	ADM29350	Adm29350 Human nov
44	32	84.2	877	4	ABG28379	Abg28379 Novel hum
45	32	84.2	1033	4	AAE07513	Aae07513 Human nuc

ALIGNMENTS

RESULT 1	
AAB02891	
ID	AAB02891 standard; peptide; 7 AA.
XX	
AC	AAB02891;
DT	18-SEP-2000 (first entry)
XX	
DE	Nucleotide-binding zinc finger alpha helix peptide, SEQ ID NO:32.
XX	
KW	zinc finger domain; alpha helix; nucleotide binding; DNA binding;
KW	polylacetyl protein; asymmetric target recognition;
KW	gene specific transcriptional regulator; gene activator; gene repressor;
KW	transcriptional switch; oncogene; erbB-2; cancer; tumour; gene therapy;
KW	transgenic animal; antiviral; anticancer; diagnosis.
OS	Synthetic.
XX	
PN	WO200023464-A2.
XX	
PD	27-APR-2000.
XX	
PF	14-OCT-1999; 99WO-EP007742.
XX	
PR	16-OCT-1998; 98US-00173941.
XX	
PA	(NOVS) NOVARTIS AG.
PA	(NOVS) NOVARTIS-RRFINDUNGEN VERW GES MBH.
PA	(SCRI) SCRIPPS RES INST.
PI	Barbas CF;
XX	
DR	WPI; 2000-339648/29.
XX	
PT	Novel isolated and purified zinc finger nucleotide-binding proteins with
PT	specificity for GNN triplet sequences, useful in gene therapy and for
PT	regulating gene function.
XX	
PS	Disclosure; Fig 1; 48pp; English.
XX	
CC	The invention relates to zinc finger nucleotide-binding proteins which
CC	comprise 2-12, preferably 2-6, operatively linked motifs selected from
CC	sequences AAB02860-B02875. Sequences AAB02860-B02875 represent the alpha
CC	helical regions of zinc finger domains which specifically bind to target
CC	nucleotide triplets of the sequence 5'-GNN-3'. Such regions may be linked
CC	by the peptide linker TGRKP (AAB02870). The Cys2-His2 zinc finger motif
CC	is the most frequently utilised nucleic acid binding motif in eukaryotes,
CC	and constitutes a beta-beta-alpha fold. Nucleic acid recognition is